Heritability and missing heritability
can twin studies be trusted?

Trzaskowski, Maciej

Awarding institution:
King's College London

HERITABILITY AND MISSING HERITABILITY: CAN TWIN STUDIES BE TRUSTED?

MACIEJ TRZASKOWSKI

MRC, SOCIAL GENETIC AND DEVELOPMENTAL PSYCHIATRY CENTRE, KING'S COLLEGE LONDON, UNIVERSITY OF LONDON

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN GENETIC RESEARCH

JUNE 2013

ABSTRACT

'Missing heritability' is the discrepancy between the amount of variance explained by specific single nucleotide polymorphisms (SNPs) identified in genome-wide association (GWA) and twin-estimated heritability. Four categories of explanations have been proposed for missing heritability: (1) additive effect sizes of common single nucleotide polymorphisms (SNPs) used in GWAS are too small to detect with current sample sizes; (2) rare variants are not captured by commercial arrays; (3) nonadditive effects (allelic, gene-gene or gene-environment interactions); (4) twin estimates of heritability are inflated. A recently developed quantitative method that uses GWA data – Genome-wide Complex Trait Analysis (GCTA) – has made it possible to explore these issues as it allows to compare quantitative twin-based estimates with quantitative DNA-based estimates. I use data from an on-going longitudinal study of ~14,000 twins (7000 pairs) born in the UK between 1994 and 1996 called the Twins Early Development Study (TEDS) to investigate the following: the proportion of twin heritability that can be explained by additive effects of common SNPs (Chapters 2, 3 and 4); increasing heritability across development in the presence of strong genetic stability (Chapters 5 and 6); and genetic pleiotropy (Chapter 7). In Chapters 2, 3 and 4, I apply univariate twin, GWA and GCTA methods to demonstrate that although we are still far from closing the gap between heritability and the actual genetic variants, there still is scope for discovery of common additive genetic effects. In Chapters 5, 6 and 7, I employ bivariate GCTA, polygenic predictor scores (PGS) and twin estimates from the same sample to confirm that twin estimates and DNA estimates of genetic pleiotropy and stability concur. In conclusion, in this thesis I provide evidence that much of the so-called 'missing heritability' can be explained by common additive genetic effects and that phenomena from twin research can be replicated using DNA alone.

ACKNOWLEDGEMENTS

My PhD programme was funded by a Medical Research Council PhD Studentship to the MRC Social, Genetic and Developmental Psychiatry Centre at the Institute of Psychiatry, King's College London. This thesis uses twin and molecular genetic data from the Twins Early Development Study (TEDS) supported by a program grant to Professor Robert Plomin from the UK Medical Research Council [G0901245, and previously G0500079], with additional support from the US National Institute of Health [HD044454; HD046167]. Genome-wide genotyping was made possible by a grant from the Wellcome Trust to the Wellcome Trust Case Consortium 2 project [085475/B/08/Z; 08475/Z/08/Z].

I am enormously grateful to my two supervisors Professor Robert Plomin and Professor Thalia C Eley for their encouragement, guidance, friendship and continued support throughout my PhD. I would also like to express my appreciation to the staff involved in the TEDS project at the Institute of Psychiatry, especially Neil Harvey for his support with many administrative obstacles, data manager Andrew McMillan for his continuous help with data extraction, and Rachel Ogden for her incredible ability to maintain TEDS. I would like to thank my friends and colleagues Dr Ken Hanscombe and Dr Rosalind Arden for their stimulating theoretical and methodological discussions. I would very much like to thank to my collaborators Dr Clare Llewellyn and Professor Jane Wardle without whom some of these projects may have not come to fruition.

I would like to express my appreciation to all the members of the MRC Social, Genetic and Developmental Psychiatry Centre who provided both invaluable advice and friendship throughout my PhD. I am eternally grateful to my family and in particular my mother whose brave decision to put my wellbeing ahead of her own resulted in this achievement.

AUTHOR DECLARATION

The work presented in this thesis was undertaken as part of the Twin Early Development Study (TEDS), an on-going longitudinal study following twins born in England and Wales between 1994 and 1996. On commencing my PhD genomic and phenotypic data were already collected. I was part of a team responsible for quality control of the molecular genetic data and responsible for preparing both phenotypic and genomic data for all analyses incorporated into this thesis. In all other respects, to the best of my knowledge, the work presented in this thesis is original and my own work, except where acknowledged in the text.

Maciej Trzaskowski

INDEX

INDEX OF TABLES

INDEX OF FIGURES

INDEX OF EQUATIONS

CHAPTER 1: INTRODUCTION

OVERVIEW

Recent advances in technology and clever implementation of statistical models have revolutionised our understanding of the genetic basis of all aspects of human life. Historically, quantitative and molecular genetics were seen as unrelated and even contradictory approaches. In molecular genetics, Mendel's law of segregation culminated in huge success in discovering genetic associations with monogenic disorders mainly through implementation of linkage designs using just a few hundred DNA markers. Since monogenic disorders are rare, the focus needed to shift to complex traits. Unfortunately, linkage designs lacked power when applied to complex traits, halting progress until the completion of the Human Genome Project, and the advent of DNA arrays, which genotyped hundreds of thousands of DNA markers throughout the genome. These two events brought fresh hope for greater insights into the genetic architecture of complex traits by enabling more precise estimation of genetic variants and more thorough coverage of the genome.

Mendel was fortunate in that he chose a monogenic trait for which underlying genetic variants not only travelled across generations according to predictions based on Mendel's first law of segregation, but also assorted independently from variants of other traits according to Mendel's second law. Unfortunately, complex traits did not conform to these laws, prompting some researchers to argue that Mendel's findings were limited to peas, and not 'higher-order' species. Although the dispute between the two factions was at times fierce, it was resolved when it became apparent that in complex traits Mendel's laws hold for segregation of single alleles but are masked by the additive influence of many such alleles. The overall genetic influence is driven by many genes, called *polygenic,* and thus the effects of individual genes is small and their overall effect on the phenotype is complex and normally distributed in a population. This is the cornerstone of quantitative genetics expounded by Fisher (1918), who described how Mendel's model could be extended to multiple genes in order to account for inheritance of complex traits. Today's methods

not only bring these two domains closer than ever before, but also show how mutually informative they are and how this reciprocity paves the way for future insights.

The current investigation begins with the problem of 'missing heritability', which started when genome-wide association (GWA) studies, although successful in identifying significant associations with complex traits, repeatedly reported significant associations that collectively accounted for only a small fraction of heritability as estimated by twin studies. The large discrepancy between twin-estimated heritability and the total variance explained by the few genes reliably associated with complex traits led to several hypotheses, one of which was that twin models have overestimated heritability. A recently developed quantitative genetic method that takes advantage of GWA data, called *Genome-wide Complex Trait Analysis* (GCTA), was first applied to the complex trait height, and showed that common additive effects can account for about half of the heritability as estimated by twin studies (Yang et al, 2010). The value of GCTA is that it is the first quantitative method based on genome-wide DNA data alone. In addition, because it uses unrelated individuals, GCTA also bypasses the equal environments assumption of the twin method (Plomin, DeFries, Knopik, & Neiderhiser, 2013). To make the best direct comparison between the molecular genetic and twin methods, I needed a sample with prospective, longitudinal data on a large number of complex traits, available from twins with genome-wide genotyping. The study that met all these criteria was the Twins Early Development Study (TEDS; Haworth, Davis, & Plomin, 2013).

Chapter 2 explores the issue of missing heritability by applying GWA and GCTA to parent-rated anxiety-related behaviours in childhood and comparing those genetic estimates to twin-estimated heritability for these traits. To check that our findings are not specific to anxiety-related behaviours or to parent ratings, I apply GCTA to other behaviour problems, and other raters (Chapter 3). Next, I turn my attention to a phenotype for which GWA has been more successful, body mass index (BMI). Meta-GWA analyses have identified 32 loci explaining about 2% of the variance in BMI, although this is still only a small fraction of twin heritability (typically about 50-90%) (Elks, den Hoed, Zhao, Sharp, Wareham, & Loos, 2012). I use GCTA to estimate the amount of BMI variance that can be explained by additive effects of all common SNPs (Chapter 4).

GCTA heritability estimates only include common additive effects of SNPs captured by current commercial DNA arrays. In contrast, twin heritability captures the aggregate effect of all genetic influence across the genome, including rarer variants, non-additive genetic effects (dominance and epistasis) and the interplay between genes and environments (interaction and correlation). Carefully comparing GCTA and twin heritability estimates can indicate the breadth and depth of the missing heritability gap by indicating the extent to which additive effects of common SNPs can account for heritability. Another approach to test the reliability of twin estimates is to focus on genetic phenomena that are not affected by the proportion of the genome that is captured. For example, one consistent finding from twin research is that heritability increases during development (e.g., Bergen et al, 2007). Age-to-age increases in heritability are well suited to direct comparison between twin and GCTA models because the proportion of genetic variance captured by commercial arrays is the same for all ages. Twin heritability also captures the same net effect of genetic influence across ages and measures. Consequently, although estimates of heritability will differ between twin and GCTA models, the trend of increasing heritability should be preserved.

Two other findings from twin studies can also be tested using GCTA: strong age-to-age genetic correlations (same genes influencing a trait at different time points, i.e. genetic stability) and genetic correlations across different phenotypes (same genes influencing different traits, also known as pleiotropy). In GCTA, genetic correlations are unbiased in that the proportion of the GCTA underestimation is the same for variance and covariance. Given that the genetic correlation is a proportion between covariance and the product of standard deviations (variance) of the two traits, it should be unaffected by this bias. For this reason, genetic correlations from age to age and across traits should be of similar magnitude in GCTA and twin studies.

Thus developmental increases in heritability, age-to-age genetic correlations, and pleiotropy are three major, consistent findings about cognitive development that are good candidates for comparing genetic correlations using GCTA and twin methods. Several phenotypes show increasing heritability in the presence of strong genetic stability (high age-to-age genetic correlations), but two lend themselves perfectly to our investigation, intelligence and BMI.

In Chapter 4 I investigate the problem of missing heritability in childhood obesity; in Chapter 5 I apply two molecular methods - bivariate GCTA and polygenic predictors - to the same phenotype, to test the twin-reported increasing heritability in the face of strong age-to-age genetic stability. Although polygenic predictors were available for BMI but not 'g', I test the same hypothesis on 'g' but the analysis is limited to bivariate GCTA only (Chapter 6). Finally, in Chapter 7, I test the pleiotropy findings from twin studies by comparing bivariate GCTA results to twin results for 'g', reading, maths and language.

METHODS

THE TWIN METHOD

For decades quantitative geneticists used family, twin and adoption designs to estimate genetic and environmental influences on many psychological and psychiatric traits. The strength of these designs came from the fact that although none was perfect, each had different limitations, and nonetheless converged on similar estimates. For example, the adoption design, which typically compares adopted children with their biological and adoptive parents, is limited by possible pre-natal factors and selective placement. In contrast, the twin design, which compares monozygotic (MZ) twins who are genetically 100% identical, and dizygotic (DZ) twins who share on average 50% of their segregating alleles, is limited by the assumption of equal environments for the two types of twins. These limitations are specific to each design and yet adoption and twin studies generally produce similar estimates of genetic and environmental influences especially for cognitive traits (Plomin et al., 2013). Adoption studies are less common today than twin studies due to the sharp reduction in neonatal adoption. For this reason the rest of this section focuses on the twin method, including its contributions and limitations.

The classical twin method takes advantage of known genetic differences between two naturally existing types of twins: MZ and DZ. Relative differences in within-pair correlations can be used to estimate the influence of additive genetic, shared environmental and unique environmental influences. Additive genetic influence refers to the effects of alleles that cumulatively add up in their effect. Shared environment refers to environmental effects that make two children in the family more similar, whereas non-shared (unique)

environments are those that lead to differences between family members. If the MZ correlation is higher than the DZ correlation for a phenotype, genetic influence (A) is assumed to play a role. Any effects that contribute to within-pair similarity but are not genetic, are attributed to shared environmental influences (C), and the extent to which MZ twins differ is attributed to non-shared environmental influences (E). E also includes random measurement error. A detailed description of this model and discussion of related issues can be found in Plomin et al. (2013).

Univariate twin studies have demonstrated consistently that virtually every human trait is influenced, at least partially, by genetic variation. Multivariate models go beyond this rudimentary finding about the relative influence of nature and nurture on a single phenotype, to investigate the common genetic architecture underlying multiple complex traits. Multivariate twin designs decompose the covariance between multiple phenotypes to provide insights into the common genetic and environmental relationship underlying different traits, as well as the genetic and environmental contributions to their development over time (trait stability and change). Multivariate genetic research has yielded three major findings: heritability increases across development, genes are developmentally stable - strong age-to-age genetic correlations – and are highly pleiotropic. Genetic pleiotropy is inferred from consistent findings of strong genetic covariance between different traits at one measurement occasion: this is true for anxiety and depression (Eley & Stevenson, 1999; Kendler, Prescott, Myers, & Neale, 2003; Thapar & McGuffin, 1997), and IQ and learning abilities (Davis, Haworth, & Plomin, 2009; Chapter 5). Genetic stability is inferred from high genetic correlations between the 'same' trait across time (Trzaskowski, Zavos, Haworth, Plomin, & Eley, 2012). Curiously, despite the observed genetic stability from age to age, many traits appear to show increases in heritability across development (Bergen, Gardner, & Kendler, 2007; Chapters 6 & 7). Although genetic stability and increasing heritability might appear paradoxical, both phenomena can be explained by gene-environment correlation, declining influence of shared environment, or new genes coming online. Unfortunately, twin designs cannot easily tease out which of these mechanisms are responsible for the phenomena. The most important advance in the field that will help to resolve this issue and many others will come when specific genetic variants are identified that account for these genetic effects. A

technological advance that greatly improved attempts to find genes responsible for heritability of complex traits is genome-wide association (GWA).

GENOME-WIDE ASSOCIATION ANALYSIS

The discovery that many complex traits are highly heritable motivated researchers to search for the genetic variants associated with these traits. Methods such as linkage analysis of DNA samples from multigenerational families with affected and unaffected members successfully identified genetic loci for many Mendelian disorders in which a single mutation is necessary and sufficient for the disease to develop. However, linkage failed to tell us much about the genes affecting complex traits (Glazier, Nadeau, & Aitman, 2002), primarily because linkage lacks power to detect small effect sizes (Risch & Merikangas, 1996) and the need for multigenerational families (pedigrees) limited potential sample sizes. Another factor was that some common disorders, such as Alzheimer's or Parkinson's disease, appear later in life. As a consequence the family pedigrees could not be constructed because parents of affected individuals would most likely have died and their offspring would be too young to reveal any symptoms (Hingorani, Shah, Kumari, Sofat, & Smeeth, 2010).

It was not until the early 21st century that two major developments finally provided cost effective, thorough coverage of the genome. The first event was the completion of the Human Genome Project (H.G.P., 2001), which resulted in the first detailed maps of the human genome and the patterns of linkage disequilibrium (LD) for hundreds of thousands SNPs. The second event was the production of DNA arrays or the 'chips'. A single genome chip could assess hundreds of thousands of target probes. Each probe is a short DNA sequence containing a common single base variant. Understanding linkage disequilibrium patterns across the genome was crucial because it showed that careful selection of a few hundreds of thousands of DNA markers was sufficient to comprehensively tag the genome in association studies. DNA variants close together on a chromosome violate Mendel's second law by segregating together producing high between-SNP correlations. A single marker from DNA variants in high LD was sufficient to 'tag' the region. This knowledge was used to select probes for genetic chips. In addition, coverage of the genome could be

improved by imputing additional SNPs from existing reference maps, such as HapMap (Gibbs et al., 2003) or, more recently, the 1000 Genomes Project (Siva, 2008). This meant that it was now possible to interrogate the whole genome simultaneously for association between allele frequency of individual SNPs and any trait varying in unrelated individuals. The method is known today as a Genome-Wide Association Study (GWAS; e.g., Cardon & Bell, 2001; Hirschhorn & Daly, 2005; Balding, 2006).

In GWA, linear regression is used to test for associations between SNPs and a quantitative trait; associations with a binary trait (e.g. disease status) employ logistic regression. Associations are usually examined one SNP at the time, assuming an additive model, although other models such as dominance can also be implemented. In the additive model, homozygotes for the increasing allele are predicted to score higher than heterozygotes and heterozygotes to score higher than homozygotes of the other allele. Although the statistical design of GWA is simple, methodological challenges are great such as data preparation and quality control (QC). A detailed explanation of all the steps involved in QC is provided in Appendix 1, but it is important to note that good QC is essential. With such a large number of multiple tests the correction for multiple testing is daunting, and the number of possible false positives is large; the accepted P-value threshold for genome-wide significance is $p<5x10^{-8}$, which can be thought of as $p<0.05$ with a Bonferroni correction for one million statistical tests. Despite such stringent correction, within the three years between 2005 and 2008, more than 400 SNP associations for just over 120 traits (www.genome.gov/GWAstudies) achieved this stringent genome-wide threshold, and revealed unexpected insights about genetic influences on complex traits (Visscher, Brown, McCarthy, & Yang, 2012). For example, in Crohn's disease, many SNPs reported through GWA studies lie in and around genes involved with autophagy, the cell's maintenance process that breaks down dysfunctional components of the cell (WTCCC, 2007). In addition, the same study showed that type 2 diabetes was associated with loci encoding for proteins relevant to insulin secretion, and not insulin signalling, as previously thought. GWA studies also provided modest molecular evidence for genetic pleiotropy – that is, the same genes affecting different traits. For example, SNPs near TCP2 gene on chromosome 8 were associated with the risk of developing type 2 diabetes as well as prostate cancer (Hingorani, et al., 2010). A coding region in the PTPN22 gene was

shown to increase risk of Type 1 diabetes and rheumatoid arthritis. There also was a strong overlap between genes involved in Crohn's disease, ulcerative colitis, ankylosing spondylitis and psoriasis (Visscher, et al., 2012).

In general, the question 'did GWA studies work?' is still unsettled. In aggregate GWA studies explain, at best, a small fraction of any trait's total variance and many fail to replicate. Small effect sizes of the significant associations suggest lack of power. But, as noted by Visscher and colleagues (2012), GWA studies have provided insight into unexpected biological pathways and mechanisms, as well as demonstrating the existence of pleiotropy.

POLYGENIC PREDICTORS

As mentioned in the previous section, one of the main reasons why GWA findings were limited was lack of power. Increasing numbers of publications consistently indicated effect sizes much smaller than originally expected. Unfortunately, the sample sizes needed to detect most of those were unattainable by individual centres, which led to the sudden emergence of world-wide consortia. This new trend provided much needed progress: GWA discoveries from 2008 (see previous section) now increased to just under 9000 SNPs in more than 700 traits (www.genome.gov/GWAstudies/). Nonetheless, the yield was still not as substantial as expected. However, in cases where associations had been detected, the SNPs could be accumulated into a polygenic score and in the aggregate could increase the total amount of variance explained. For example, if we had 10 individual SNPs, each with an effect size of .5%, a polygenic score comprising all of these would account for 5% of the variance giving a sample of 150 individuals 80% power to detect their cumulative effect.

The creation of a polygenic score usually concerns three aspects of the association: the direction, the magnitude and the significance. Given that all the 'candidate' SNPs are 'pooled' together into a single score, the direction of their effect has to be consistent. These can then be added up into a single score (although summing scores is most commonly used, non-additivity can also be incorporated). To give priority to stronger associations, the scores can be weighted by the betas from the regression (i.e., their effect size).

Finally, the SNPs can either be selected from previously reported significant 'hits' or they can be amassed from a single GWAS by selecting everything above an arbitrary p-value. The former method is preferred but the choice is, of course, dependent on the availability of previously reported 'robust hits' because of the trade-off between specificity and size. The tiny effect sizes and moderate-to-large estimates of heritability suggest that the more SNPs you 'pool' together the more variance you should explain. However, adding large number of SNPs with no or opposite effect can easily 'dilute' or 'mask' the signal.

BMI is an example of a successful GWA that resulted in polygenic analysis (Speliotes et al., 2010). The authors used a set of meta- and joint analyses to replicate earlier reported genome-wide significant hits and to discover an additional 22 SNPs, resulting in a set of 32 loci robustly associated with BMI. The authors then summed the 32 SNPs to create a polygenic score to test their cumulative effect on BMI. They reported that even though the SNPs accounted for only ~1.5% of the variance in BMI (a third of which was accounted for by the FTO gene), an individual at high genetic risk (≥38 risk alleles) versus low genetic risk (≤ 21 risk alleles) differed by 2.73 BMI units (or ~9kg for an adult of average height).

Importantly for us, one question that can be tackled using the polygenic predictors is why twin-estimated heritability increases over development, in the presence of strong age-to-age genetic correlations. As indicated in Chapters 5 and 6, it has been consistently shown that heritability of some phenotypes increases over development. Several hypotheses have been proposed to explain this phenomenon - e.g. novel genes coming 'online' at different stages of development, new genes coming online, or gene-environment correlation. Twin studies are unable to distinguish between these. Fortunately for us, heritability increases for BMI (Haworth et al., 2008), and we now also have more than 30 SNPs robustly associated with it (Speliotes, et al., 2010). We can therefore use these SNPs to test the whether the increase in heritability observed for BMI can be explained by the same set of SNPs (Chapter 7).

Genome-wide Complex Trait Analysis (GCTA)

As indicated earlier, the mismatch between the high heritability estimates derived from twin studies and the small proportion of variance explained from common variants identified through GWAS has come to be

known as the problem of 'missing heritability' (Maher, 2008). Twin studies suggest that almost every aspect of human life, from our biology through to cognition and behaviour is heritable, with genetic influences often explaining moderate (30-40%) to very high (80-90%) proportions of variance. Finding only a handful of associated genetic loci through the GWA method was disappointing and resulted in papers proposing ways to narrow the missing heritability gap (e.g. McCarthy & Hirschhorn, 2008). The most general explanation was that the influence of each individual SNP was so small that most of the GWA studies thus far were greatly underpowered to detect them. In addition, the markers selected for the genetic arrays were limited to common variants only (minor allele frequency > 1%). This meant that many potential 'true' associations with SNPs of lower allele frequency would be missed due to low LD with the markers. It was also likely that non-additive effects, such as gene-environment or gene-gene interactions, were of greater importance than originally assumed – GWA is limited to additive genetic effects. Finally, there was a widespread scepticism about the reliability of heritability estimates derived from twin data, implying that twin estimates were inflated.

A recently proposed method that uses genome-wide genotyping data in a novel way has potential for answering some of these questions about the genetic architecture of complex traits and common disorders. The method is incorporated in the Genome-wide Complex Trait Analysis (GCTA) package which provides a pipeline for estimating the amount of phenotypic variance (Yang, Lee, Goddard, & Visscher, 2011) and co-variance (Lee, Yang, Goddard, Visscher, & Wray, 2012) that is explained by all SNPs available on a DNA array and all DNA variants in LD with them. The method requires genome-wide data for unrelated individuals. It obtains unbiased estimates of heritability using restricted maximum likelihood (REML). The variance (or covariance) of a continuous trait is 'partitioned' using a mixed linear model (MLM), and in the case of a binary trait, using a liability threshold model. Given that this thesis discusses continuous traits only I will limit the discussion to MLM only.

Before the variance or covariance can be decomposed into genetic and residual components, pairwise genomic similarity is calculated between all pairs of individuals in the sample using all genetic markers genotyped on the SNP array. Specifically, genotypic similarity between two individuals is weighted by allele

frequencies at each SNP, and the final genetic relatedness coefficient is a mean of all these scores (see equation 1; adapted from Yang, Lee, et al., 2011). Of note, the algorithm is designed to give precedence to rarer alleles, with assumption that sharing rarer alleles will be more important to phenotypic similarity. Because the GCTA package uses a random effects model to estimate genetic effects from a sample of unrelated individuals in the population, any pair whose genetic similarity is equal to or greater than a fourth cousin is removed, which represents an estimate of pairwise relatedness > 0.025.

Equation 1. Algorithm for the estimation of pairwise genetic similarity

$$\hat{\pi}_{ij} = \frac{1}{N}\sum_k \frac{(x_{ki} - 2p_k)(x_{kj} - 2p_k)}{2p_k(1-2p_k)}$$

In univariate analysis, the variance of a trait is partitioned using residual maximum likelihood (REML) into polygenic (g) and residual effects (e). The genetic component of the decomposition is a product of genetic similarity (A) and the estimated additive genetic effect (σ^2_g - Equation 2). The model can also include fixed effect covariates, such as stratification axes or gender (not shown in the equation).

Equation 2. Univariate mixed linear model (MLM)

$$\hat{\mathrm{var}}(y) = A\hat{\sigma}^2_g + I\hat{\sigma}^2_e$$

The bivariate method extends the univariate model by relating the pairwise genetic similarity matrix to a phenotypic covariance matrix between traits 1 and 2 (Lee, et al., 2012). Here the residual component is modelled only on the variance of each trait, whereas their covariance is strictly due to fixed covariates and random genetic effects (Equation 3). A is a genetic relatedness matrix and Z is an incidence matrix for the random polygenic effect. The sampling distribution of the statistic is calculated using the average information (AI) method (Sang Hong Lee & van der Werf, 2006) and these are then used to derive the standard error of each estimate.

Equation 3. Variance / covariance matrix of bivariate mixed linear model

$$V = \begin{bmatrix} Z_1 A Z_1^{'} \sigma^2_{g_1} + I\sigma^2_{e_1} & Z_1 A Z_1^{'} \sigma^2_{g_1 g_2} \\ Z_1 A Z_1^{'} \sigma^2_{g_1 g_2} & Z_2 A Z_2^{'} \sigma^2_{g_2} + I\sigma^2_{e_2} \end{bmatrix}$$

Because GCTA is based on genome-wide DNA data alone, it can be used to estimate genetic influence for unrelated individuals rather than requiring special relatives such as MZ and DZ twins. Because it uses genetic data from measured genotypes in unrelated individuals, it bypasses some of the assumptions of the twin method (Plomin, et al., 2013). GCTA estimates how much of the phenotypic variance is explained by all additive genetic effects captured by commercial gene-chips with common SNPs. That is, it captures all of the genetic markers placed on the array and everything in LD with them. However, the strengths of GCTA are also its weaknesses: it can only detect genetic influence due to the additive effects of common SNPs that are included on currently available DNA arrays. Nonetheless, GCTA provides important information about the extent to which the genetic architecture of complex traits includes additive effects of common SNPs, and sets the limit for detecting associations in GWA studies.

GCTA analyses have shown that information captured by current DNA arrays can explain a substantial amount of the variance in several complex traits, including human height (Yang et al., 2010), BMI (Yang et al., 2011), psychiatric and medical disorders (S. H. Lee, Wray, Goddard, & Visscher, 2011; S. H. Lee, et al., 2012; Lubke et al., 2012), personality (Vinkhuyzen et al., 2012), and cognitive traits (Deary et al., 2012). However, thus far there is only one study that made a direct comparison between GCTA and the twin method using the same sample (Plomin et al., 2013). I therefore set out to perform such comparisons in a variety of developmentally important traits (Chapters 2, 3 and 4).

The bivariate extension of GCTA has also opened up the opportunity to test multivariate hypotheses. Importantly, these can be tested without the limitations of twin models. One of the most widely reported findings from twin analyses is pleiotropy (also known as the 'Generalis Genes Hypothesis'; R. Plomin & Kovas, 2005). By employing GCTA, I can test this mechanism using DNA alone. The TEDS sample allows me

to apply twin and GCTA methods to the same sample to test their agreement. A second general finding from twin studies that can be tested using bivariate GCTA is genetic stability across development – i.e. that genes largely account for age-to-age stability, despite evidence for increasing heritability.

SUMMARY

In summary, in this thesis I set out to investigate the proportion of 'missing heritability' in childhood across several complex phenotypes, including anxiety-related behaviours, ADHD, autistic-like traits, general intelligence, and learning abilities (Chapters 2, 3 and 4). I also use molecular genetic methods to test some of the most consistent findings from twin studies: increasing heritability in presence of strong genetic stability (Chapters 5 and 6), and strong pleiotropy (Chapter 7).

REFERENCES

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. [Research Support, Non-U.S. Gov't Review]. *Nature reviews. Genetics, 7*(10), 781-791. doi: 10.1038/nrg1916

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Research and Human Genetics, 10*(3), 423-433. doi: 10.1375/twin.10.3.423

Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review]. *Nature reviews. Genetics, 2*(2), 91-99. doi: 10.1038/35052543

Davis, O. S. P., Haworth, C. M. A., & Plomin, R. (2009). Learning abilities and disabilities: Generalist genes in early adolescence. *Cognitive Neuropsychiatry, 14*, 312-331. doi: 10.1080/13546800902797106

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D., . . . Corley, J. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482*(7384), 212-215.

Eley, T. C., & Stevenson, J. (1999). Exploring the covariation between anxiety and depression symptoms: a genetic analysis of the effects of age and sex. *J Child Psychol Psychiatry., 40*(8), 1273-1282.

Elks, C. E., den Hoed, M., Zhao, J. H., Sharp, S. J., Wareham, N. J., Loos, R. J., & Ong, K. K. (2012). Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne),* 3, 29. doi: 10.3389/fendo.2012.00029

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*, 399-433.

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., . . . Shen, Y. (2003). The international HapMap project. *Nature, 426*(6968), 789-796.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science, 298*(5602), 2345-2349.

H.G.P. (2001). The Human Genome. *Nature, 409*(6822), 745-964.

Haworth, C., Carnell, S., Meaburn, E. L., Davis, O. S. P., Plomin, R., & Wardle, J. (2008). Increasing heritability of BMI and stronger associations with the FTO gene over childhood. *Obesity, 16*(12), 2663-2668.

Haworth, C. M., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): A Genetically Sensitive Investigation of Cognitive and Behavioral Development From Childhood to Young Adulthood. *Twin Res Hum Genet, 16*(1), 117-125. doi: 10.1017/thg.2012.91

Hingorani, A. D., Shah, T., Kumari, M., Sofat, R., & Smeeth, L. (2010). Translating genomics into improved healthcare. *BMJ, 341*(nov05 1), c5945-c5945. doi: 10.1136/bmj.c5945

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. [10.1038/nrg1521]. *Nat Rev Genet, 6*(2), 95-108.

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry., 60*(9), 929-937.

Lee, S. H., & van der Werf, J. H. J. (2006). An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genetics Selection Evolution, 38*(1), 1-19.

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. [Research Support, Non-U.S. Gov't]. *American Journal of Human Genetics, 88*(3), 294-305. doi: 10.1016/j.ajhg.2011.02.002

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. doi: 10.1093/bioinformatics/bts474

Lubke, G. H., Hottenga, J. J., Walters, R., Laurin, C., de Geus, E. J., Willemsen, G., . . . Boomsma, D. I. (2012). Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol Psychiatry, 72*(8), 707-709. doi: 10.1016/j.biopsych.2012.03.011

Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature, 456*, 18-21. doi: 10.1038/456018a

McCarthy, M. I., & Hirschhorn, J. N. (2008). Genome-wide association studies: potential next steps on a genetic journey. *Human molecular genetics, 17*(R2), R156-165. doi: 10.1093/hmg/ddn289

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral Genetics* (6 ed.). New York: Worth Publishers.

Plomin, R., Haworth, C. M., Meaburn, E. L., Price, T. S., & Davis, O. S. (2013). Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychological Science, 24*(4), 562-568. doi: 10.1177/0956797612457952

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131*, 592-617.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science, 273*(5281), 1516-1517.

Siva, N. (2008). 1000 Genomes project. *Nature biotechnology, 26*(3), 256-256.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., . . . MV§gi, R. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics, 42*(11), 937-948.

Thapar, A., & McGuffin, P. (1997). Anxiety and depressive symptoms in childhood - A genetic study of comorbidity. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*(6), 651-656.

Trzaskowski, M., Zavos, H. M., Haworth, C. M., Plomin, R., & Eley, T. C. (2012). Stable genetic influence on anxiety-related behaviours across middle childhood. *J Abnorm Child Psychol., 40*(1), 85-94.

Vinkhuyzen, A. A., Pedersen, N. L., Yang, J., Lee, S. H., Magnusson, P. K., Iacono, W. G., . . . Wray, N. R. (2012). Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Translational psychiatry, 2*, e102. doi: 10.1038/tp.2012.27

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five Years of GWAS Discovery. *The American Journal of Human Genetics, 90*(1), 7-24. doi: 10.1016/j.ajhg.2011.11.029

WTCCC. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*(7145), 661-678. doi: 10.1038/nature05911

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S. D., Henders, A. K., Nyholt, D. R., . . . Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. [10.1038/ng.608]. *Nature Genetics, 42*(7), 565-569. doi: http://www.nature.com/ng/journal/v42/n7/suppinfo/ng.608_S1.html

Yang, J., Lee, S., Goddard, M., & Visscher, P. (2011). GCTA: A tool for genome-wide complex trait analysis. [Article]. *American Journal of Human Genetics, 88*(1), 76-82. doi: 10.1016/j.ajhg.2010.11.011

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., . . . Hayes, M. G. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics, 43*(6), 519-525.

CHAPTER 2: FIRST GENOME-WIDE ASSOCIATION STUDY ON ANXIETY-RELATED BEHAVIOURS IN CHILDHOOD

This chapter is presented as a published paper and is an exact copy of the following journal publication:

Maciej Trzaskowski*, Thalia C. Eley*, Oliver S.P. Davis, Sophia J. Doherty, Ken B. Hanscombe, Emma L. Meaburn, Claire M.A. Haworth, Thomas Price, Robert Plomin (2013). *PloS one, 8*(4), e58676. doi: 10.1371/journal.pone.0058676

* These authors contributed equally

OPEN ACCESS Freely available online

PLOS ONE

First Genome-Wide Association Study on Anxiety-Related Behaviours in Childhood

Maciej Trzaskowski[1]*, Thalia C. Eley[1], Oliver S. P. Davis[1], Sophia J. Doherty[1], Ken B. Hanscombe[1], Emma L. Meaburn[2], Claire M. A. Haworth[1], Thomas Price[1], Robert Plomin[1]

1 MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, De Crespigny Park, London, United Kingdom, **2** Birkbeck, Department of Psychological Sciences, University of London, London, United Kingdom

Abstract

Background: Twin studies have shown that anxiety in a general population sample of children involves both domain-general and trait-specific genetic effects. For this reason, in an attempt to identify genes responsible for these effects, we investigated domain-general and trait-specific genetic associations in the first genome-wide association (GWA) study on anxiety-related behaviours (ARBs) in childhood.

Methods: The sample included 2810 7-year-olds drawn from the Twins Early Development Study (TEDS) with data available for parent-rated anxiety and genome-wide DNA markers. The measure was the Anxiety-Related Behaviours Questionnaire (ARBQ), which assesses four anxiety traits and also yields a general anxiety composite. Affymetrix GeneChip 6.0 DNA arrays were used to genotype nearly 700,000 single-nucleotide polymorphisms (SNPs), and IMPUTE v2 was used to impute more than 1 million SNPs. Several GWA associations from this discovery sample were followed up in another TEDS sample of 4804 children. In addition, Genome-wide Complex Trait Analysis (GCTA) was used on the discovery sample, to estimate the total amount of variance in ARBs that can be accounted for by SNPs on the array.

Results: No SNP associations met the demanding criterion of genome-wide significance that corrects for multiple testing across the genome ($p<5\times10^{-8}$). Attempts to replicate the top associations did not yield significant results. In contrast to the substantial twin study estimates of heritability which ranged from 0.50 (0.03) to 0.61 (0.01), the GCTA estimates of phenotypic variance accounted for by the SNPs were much lower 0.01 (0.11) to 0.19 (0.12).

Conclusions: Taken together, these GWAS and GCTA results suggest that anxiety – similar to height, weight and intelligence – is affected by many genetic variants of small effect, but unlike these other prototypical polygenic traits, genetic influence on anxiety is not well tagged by common SNPs.

Citation: Trzaskowski M, Eley TC, Davis OSP, Doherty SJ, Hanscombe KB, et al. (2013) First Genome-Wide Association Study on Anxiety-Related Behaviours in Childhood. PLoS ONE 8(4): e58676. doi:10.1371/journal.pone.0058676

Editor: Abraham A. Palmer, University of Chicago, United States of America

Received October 30, 2012; **Accepted** February 5, 2013; **Published** April 2, 2013

Funding: TEDS is supported by a program grant (G0500079) from the U.K. Medical Research Council (MRC), with additional support from the US National Institutes of Health (HD044454; HD046167). Genome-wide genotyping was made possible by grants from the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z; 085475/Z/08/Z). RP is supported by a Medical Research Council Research Professorship award (G19/2) and a European Advanced Investigator award (295366); O.S.P.D. is supported by a Sir Henry Wellcome Fellowship from the Wellcome Trust (WT088984); C.M.A.H. is supported by a research fellowship from the British Academy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: maciej.trzaskowski@kcl.ac.uk

These authors contributed equally.

Introduction

Anxiety disorders are among the most common psychiatric disorders [1]. They often begin in childhood [2] and continue into adulthood [3], when they become co-morbid with other psychiatric disorders especially depression [4] and entail significant costs both to society and to the individual [5]. Quantitative anxiety-related traits, assessed as clinical symptoms, e.g. [6] or personality/temperament traits [7,8], are strong predictors of diagnosed anxiety disorders [7].

Twin studies have shown that childhood anxiety in representative samples, like other complex traits, is influenced genetically, e.g. [9]. Multivariate genetic studies indicate genetic overlap as well as specificity between different aspects of anxiety and from age to age as early as the preschool years [10] and into middle childhood [11] and adolescence [12,13]. At age 7, the age of the twins in the present study, parent ratings of anxiety-related traits have been shown to be moderately heritable with both domain-general and trait-specific genetic effects [11]. Similar results were found at age 9 and for continuity from age 7 to age 9 [14]. Although these quantitative genetic findings are important, the next step is to identify specific genes responsible for these effects.

Until recently, molecular genetic investigation into the aetiology of anxiety relied on linkage and candidate-gene designs. Linkage, which looks for co-inheritance between DNA variants and a disorder within families, is a systematic strategy for detecting genes of large effect size throughout the genome. However, linkage

found few such large effects for common disorders like anxiety and lacks power to detect more modest effects [15].

In contrast, allelic association, which looks for correlations between an allele and a trait among unrelated individuals, is much more powerful than linkage, but until recently, association has been limited to the exploration of a few candidate genes and could not be used to conduct a systematic search of the genome. Candidate-gene association studies of anxiety-related traits reported many associations but few of these associations have stood the test of replication, similar to candidate-gene studies in other domains in the life sciences [16].

Association studies became systematic with the advent of genome-wide DNA arrays that genotype hundreds of thousands of DNA variants throughout the genome and resulted in a plethora of genome-wide association (GWA) studies [17]. Although the first major GWA studies were reported in 2007 [18], significant results have been reported for more than 200 traits in 1500 GWA studies [19]. The only GWA studies of anxiety-related traits have focused on the personality trait of neuroticism in adults and reported possible associations with several genes [20,21,22]. However, no GWA studies of anxiety-related traits in children have previously been reported.

The current study presents the first GWA study of anxiety-related traits in children. The multivariate genetic results mentioned earlier led us to consider trait-specific as well as domain-general measures. Despite the success of GWA, reported associations are of small effect size and together account for only a modest proportion of the heritability of traits, known as the "missing heritability" problem [23,24]. One of many possible reasons for the missing heritability problem is that potential associations are missed by the common SNPs that are included in extant DNA arrays. To test this hypothesis, a new technique, described by Yang et al. [25] and implemented in a software package called *Genomewide Complex Trait Analysis (GCTA)*, has been developed that allows estimation of the total genetic variance captured by SNPs on a genome-wide DNA array, even though it does not identify which SNPs are responsible for the genetic influence [26]. For this reason, we also report GCTA results for anxiety-related traits in childhood and compare them to our twin study estimates of heritability from the same sample at the same age and using the same measures.

Methods

Ethics Statement

Written parental consent was obtained prior to data collection and the project received approval from the Institute of Psychiatry ethics committee (05/Q0706/228).

Sample

The sample was drawn from the Twins Early Development Study (TEDS), a multivariate longitudinal study which recruited over 11,000 twin pairs born in England and Wales in 1994, 1995 and 1996 [27], whose families are representative of the UK population [28]. Twins with severe medical problems or severe birth complications or whose zygosity could not be determined were excluded from the sample. To decrease heterogeneity of ancestry, the sample was restricted to families who identified themselves as white and whose first language was English. After exclusions, 7834 pairs of twins had anxiety data available at age 7 (mean age = 7.06, SD = 0.25). Although anxiety data were also available at age 9, we did not use these data in our GWA analyses because only half the sample were contacted at age 9 to provide phenotypic data.

3747 DNA samples from unrelated children in TEDS were sent for DNA array genotyping at the Wellcome Trust Sanger Institute, Hinxton, UK as part of the Wellcome Trust Case Control Consortium 2.

3665 samples were successfully hybridized to Affymetrix GeneChip 6.0 SNP genotyping arrays using standard experimental protocols (see Text S1). 3152 samples (1446 males and 1706 females) survived stringent quality control procedures performed (see Text S1), of whom 2810 also had anxiety data.

The replication sample was also drawn from TEDS children for whom DNA and anxiety data were available but for whom genome-wide genotyping was not available. After quality control, both anxiety data and SNP genotyping were available for 4804 additional individuals. Of these, 2625 were unrelated children who were also unrelated to children in the discovery sample; for 1742 children, their fraternal co-twin was in the discovery sample, and for 437 children their fraternal co-twin was also in the replication sample.

Anxiety-Related Behaviours Questionnaire (ARBQ)

Anxiety was rated by parents using the Anxiety-Related Behaviours Questionnaire (ARBQ) [10]. The ARBQ is a quantitative trait parent rating instrument for children in the general population rather than a diagnostic tool. It includes items that assess anxiety symptoms as well as aspects of anxiety-related personality. The items are best structured as four latent variables in childhood: negative affect, negative cognition, fear, and social anxiety [11]. In order to investigate domain-general genetic associations, we also constructed a general anxiety composite by summing the standardised scores for these four variables. The overall composite was crucial to produce a phenotypic measure that was free from any scale-specific error. In addition, combining standardised scores assured that none of the scales biased the composite. The ARBQ has been shown to have good construct validity, and high internal consistency [10]. In order to avoid the skew that occurs for behaviour problem measures, the five anxiety scores were quantile normalised (van der Waerden; ranks averaged for tied data) [29]. Although the distributional properties of these transformed scores are better, the correlation between the raw scores and the transformed scores varied from .80 to .98 and results were highly similar for the raw and transformed scores.

Genotyping

Genome-wide genotyping was done on AffymetrixGeneChip 6.0 SNP genotyping array with additional ~2.5 million SNPs imputed from HapMap 2 and 3 and WTCCC controls Details about genotyping and quality control are included in the Text S1. 13 SNPs for the top hits for the five anxiety-related scales from the discovery sample were genotyped in the replication sample of 4804 individuals using the Sequenom MassARRAY iPlex Gold® system (Sequenom, San Diego, USA). Three SNPs failed to meet quality control criteria, leaving 10 SNPs available for the replication stage.

Statistical Analyses

Genome-wide association (GWA) analysis. Linear regression analyses were conducted using SNPTEST v2.0 [18] under an additive model, using a frequentist method that accounts for uncertainty of genotype information [30]. We included age, sex, cohort and eight eigenvectors representing population ancestry as covariates. Consolidation and summary of the GWA results was performed in R (www.r-project.org) [31].

The strongest association results from the GWA were selected for genotyping in the replication sample. Where imputed SNPs were in LD with genotyped SNPs, the genotyped SNPs were

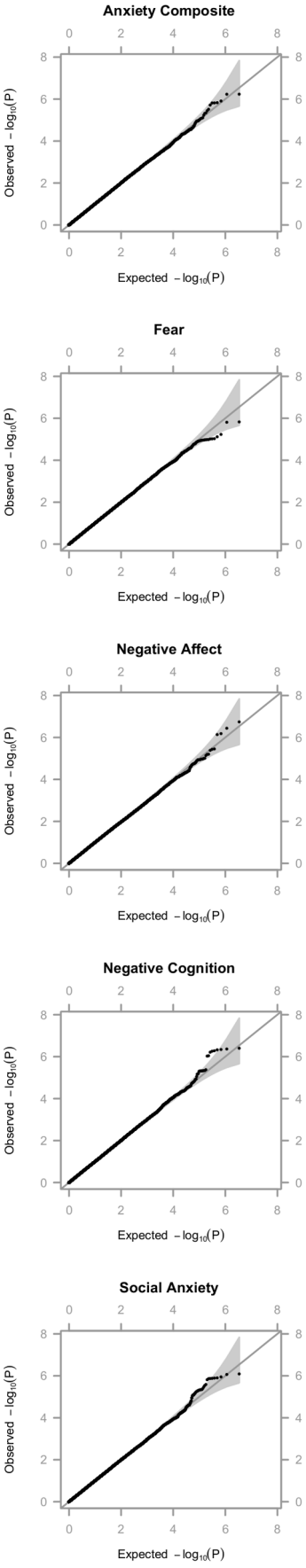


Figure 1. Log quantile-quantile (Q–Q) p-value plots for 1,724,317 single-SNP test of association of four anxiety-related traits and the anxiety composite at age 7. Footnote: Expected (X-axis) versus observed (Y-axis) p-values are plotted on the negative log scale to highlight the strongest associations. The diagonal line represents the null hypothesis and the grey polygons represent the 95% confidence interval (CI) of the null range. Significant association would be indicated by departure of the p-value (black dot) beyond the 95% CI of the null range.
doi:10.1371/journal.pone.0058676.g001

preferred. However, one especially promising imputed SNP (rs1113313) was also selected. The SNPs were selected that were not in linkage disequilibrium (LD) with each other.

Sequenom genotyping results for the replication sample were analysed using the same protocols and software as those in GWA analysis. We conducted analyses using the total replication sample as well as the subsample of individuals genetically unrelated to each other or to individuals in the discovery sample. Although this is somewhat unorthodox, power is crucial for replication and the total sample provides maximum power because it maximises sample size. If replication is found for the total sample, the replication may be biased because the sample is not completely independent of the discovery sample and more replication would be required for definitive proof of replication. However, if the results from the discovery sample do not replicate using the total sample, this is the strongest possible evidence of failure to replicate because our replication sample consists of a highly similar sample tested at exactly the same age using exactly the same measures.

Genome-wide Complex Trait Analysis (GCTA). GCTA does not attempt to identify specific variants associated with traits. Instead, it uses chance genetic similarity among unrelated individuals across hundreds of thousands of SNPs to predict phenotypic similarity. We used the GCTA software package [25] to evaluate the amount of the phenotypic variance explained by the genetic information available from the Affymetrix 6.0 DNA array. Detailed explanation of the methodology and procedure is available from Yang et al. [26]. To remain consistent with the procedure outlined by the proponents of the software, we initially used all ~700,000 genotyped SNPs to calculate a genetic relatedness matrix (GRM). However, GCTA results reported previously for height, weight and intelligence used the Illumina microarray, which was designed with specific focus on European ancestry, whereas the Affymetrix microarray was less ancestry specific. We found that by adding high-quality imputed SNPs (see 'Genotyping' section), thus increasing the number of SNPs to ~1.7 million, brought our GCTA estimates in line with previously published estimates for height, weight and intelligence. Thus, we used the 1.7 million SNPs to estimate how much of the heritability as estimated by the classical twin method could be accounted for by the available genetic information.

Results

Genome-wide Association (GWA)

Figure 1 presents quantile-quantile (Q-Q) plots for the five anxiety-related traits. Q–Q plots graphically compare the ~1.7 million observed $-\log_{10}$ p values against the $-\log_{10}$ p values expected on the basis of no association. Although there is some increase of observed p values against expected p values for the lowest p values, few of the associations fall outside the 95% confidence bands (the grey areas), which indicates that there is little evidence of significant deviation from the null hypothesis of no association.

Figure 2 presents 'Manhattan' plots for the same traits that show $-\log_{10}$ p values on the Y axis for the ~1.7 million SNPs across the 22 autosomes on the X axis. The p values on the Y axis are the negative logarithms of the p values so that the highest points in the plot represent the strongest SNP associations. The dotted horizontal line represents suggestive significance (5×10^{-7}), not genome-wide significance (5×10^{-8}). Regions with the strongest associations were chosen for replication – for example, regions of chromosome 6 and 12 that reached suggestive significance (5×10^{-7}) for the anxiety composite and negative cognition scale respectively. The association of the SNP on chromosome 6 with negative affect (Figure 2) was not proposed for replication due to the SNP's low minor allele frequency (maf = 0.03). Table 1 shows results in the discovery sample for the 10 SNPs that were also successfully genotyped in the replication sample. Two of the lowest p values in the discovery sample were SNP rs16879771, associated with the anxiety composite ($p = 6.27\times10^{-7}$), and rs1952500, which was associated with Negative Cognition ($p = 4.12\times10^{-7}$). The significance of the remaining SNPs varied from 10^{-4} to $8-10^{-7}$. The amount of variance explained in the discovery sample as indicated by the squared beta values varied from 0.09% to 1.0%. Visual inspection of the genotype-specific means suggested that none of the selected SNPs deviated from additivity.

Replication

Table 1 also includes results for the 10 SNPs in the replication sample. None of the SNPs reached significance and the direction of the associations in the replication sample was nearly at a chance level (6 in the same direction as in the GWA analysis and 4 in the opposite direction). These replication analyses were based on our total replication sample of 4804 for which we had greatest power; similarly negative results were found for our subsample of 2625 individuals which constituted a more independent but less powerful replication sample.

Genome-wide Complex Trait Analysis (GCTA)

As described earlier, we used ~1.7 million SNPs to estimate the GCTA Genetic Relatedness Matrix for our sample of 2810 individuals. Our sample included no known pairs related in the traditional sense, which was confirmed by finding that no pairs reached the standard GCTA relatedness cut-off threshold of 0.025 genetic relatedness. Table 2 summarises the GCTA estimates obtained for the five anxiety-related traits and compares them to twin study heritability estimates from the sample at the same age using the same measures. Table 2 also includes GCTA estimates for height and weight in our sample in order to compare our results to previously reported results for height and weight. As indicated in Table 2, our twin study heritability estimates are 0.80 and 0.84 for height and weight, respectively, and our GCTA estimates are 0.35 and 0.42, all of which are comparable to results reported in the literature [32]. Also similar to the literature reviewed in the Introduction, our twin study heritabilities for anxiety-related traits are substantial, varying from 0.50 to 0.61. However, the GCTA estimates for anxiety-related traits were much lower, ranging from only 0.01 to 0.19. None of the GCTA estimates reached statistical significance ($p<.05$) due to the large standard errors of estimates.

Discussion

This first genome-wide association study of anxiety-related traits in childhood indicates that no common genetic variants of large effect contribute to the heritability of these traits. Our sample of 2810 had 80% power to detect causal variants with effect sizes

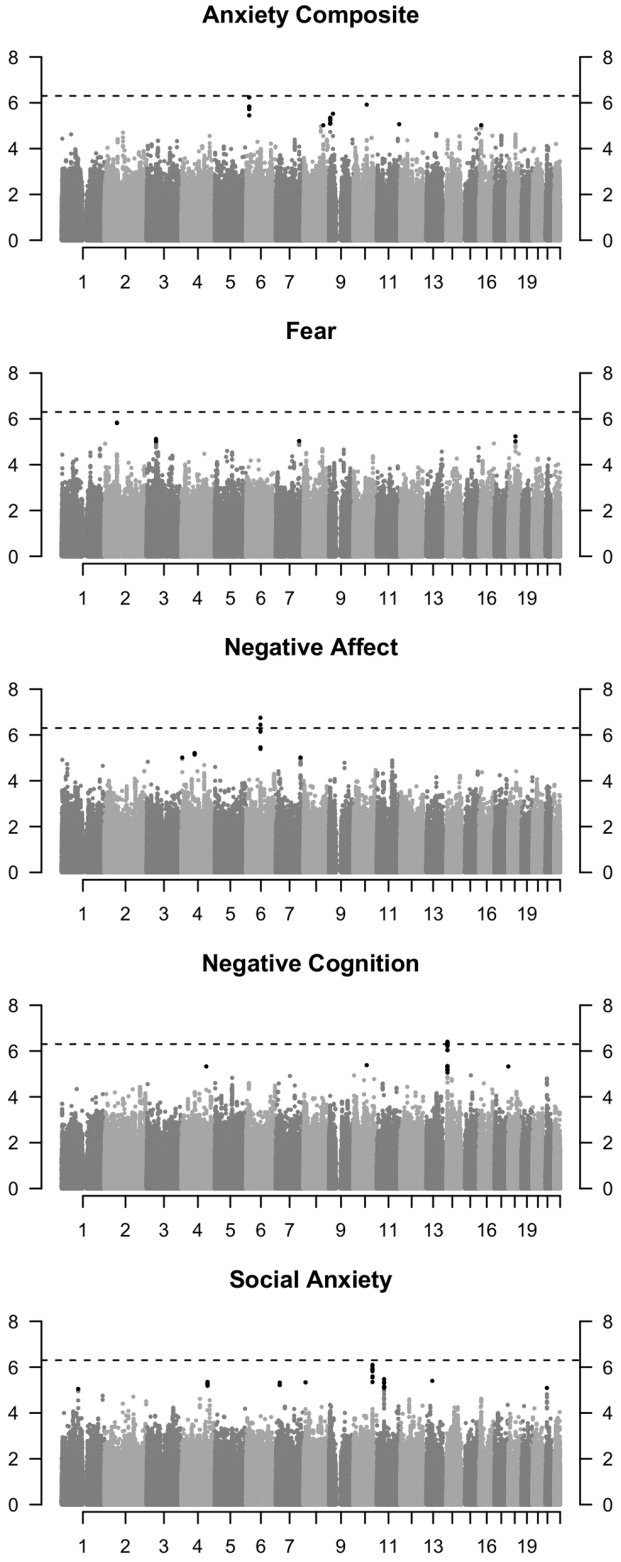


Figure 2. Manhattan plots for 1,724,317 single-SNP test of association for four anxiety-related traits and anxiety composite at age 7. Footnote: Oberved p-values are plotted on a scale of negative logs (Y-axis) against the SNP's physical position in the genome (X-axis). Black dots represent associations with $p<5\times10^{-5}$ and the horizontal dashed line represents suggestive significance with $p<5\times10^{-7}$.
doi:10.1371/journal.pone.0058676.g002

Table 1. Associations in the GWA discovery sample and in the replication sample for SNPs showing the lowest p values in the GWA analysis.

rsid	*chr*	*alleleA*	*alleleB*	*maf*	*gene*	*Phenotype*	*Beta discovery*	*Beta replication*	*P-value discovery*	*P-value replication*
rs7649323	3	C	G	0.36	DCP1A	Fear	0.06	0.02	7.42×10^{-6}	0.11
rs4568308	4	A	G	0.21	EREG, BTC, AREG	Negative Affect	0.07	−0.003	6.73×10^{-6}	0.28
rs16879771	6	C	T	0.08	CAP2	Anxiety Composite	−0.07	−0.03	6.27×10^{-7}	0.15
rs4130405	8	A	C	0.15	NIPAL2, KCNS2	Anxiety Composite	0.05	0.01	1.70×10^{-5}	0.40
rs1113313	10	C	T	0.41	VDAC2, SAMD8	Negative Cognition	−0.06	−0.04	4.20×10^{-6}	0.44
rs2772129	10	A	G	0.38	SORCS1, XPNPEP1	Social Anxiety	0.07	−0.02	8.68×10^{-7}	0.14
rs10787217	10	A	T	0.24	SORCS1, XPNPEP1	Negative Affect	−0.06	0.01	9.98×10^{-5}	0.28
rs2922037	11	C	T	0.16	API5, LRRC4C	Social Anxiety	0.08	0.02	8.22×10^{-6}	0.13
rs1952500	14	A	C	0.12	STXBP6, NOVA1	Negative Cognition	0.10	−0.06	4.12×10^{-7}	0.13
rs9977125	21	C	T	0.38	TMPRSS15, C21orf131	Anxiety Composite	0.03	0.02	1.20×10^{-4}	0.05

Footnote: rsid – SNP id; chr – chromosome; maf – minor allele frequency; gene – nearest gene; p-values for the replication sample are one-tailed and uncorrected for multiple testing.

doi:10.1371/journal.pone.0058676.t001

greater than 1.4% of the variance and none were detected. Power was calculated with the Genetic Power Calculator [33] using an additive model with a genome-wide significance threshold of $p<5\times10^{-8}$. As seen in Table 1, the largest effect size from the GWA analysis accounted for only 1% of the variance. Our power calculations indicated that we had less than 80% of power to detect a signal of this magnitude; thus this result should be considered with caution until replicated. That said, these results are similar to those found for other quantitative traits for which the strongest associations account for about 1% of the variance such as height [34], weight (GIANT Consurtium) [35], and cognitive traits including reading [36], mathematics [37], and general cognitive ability [38,39]. Our GWA results for anxiety-related traits in childhood are compatible with a growing consensus from GWA studies of complex traits that the largest effect sizes are very small and that all known associations only explain a small portion of the heritabilities of complex traits and common disorders, a gap that is known as the *missing heritability problem* [23]. The missing heritability problem can be seen in Table 2 in which our twin study estimates of heritability for the five anxiety-related scales exceed 50%, whereas the sum of the effect sizes of the 10 SNPs shown in Table 2 is less than 5% in the discovery sample, and negligible in the replication sample.

Table 2. Genome-wide Complex Trait Analysis (GCTA) estimates of genetic variance compared to twin study estimates of heritability.

Phenotype	*GCTA estimate (SE)*	*P-value*	*Twin study h^2 estimate (SE)*
Negative Cognition	0.07 (0.12)	0.257	0.52 (0.03)
Negative Affect	0.07 (0.12)	0.281	0.50 (0.03)
Fear	0.19 (0.12)	0.057	0.59 (0.02)
Social Anxiety	0.01 (0.11)	0.479	0.61 (0.01)
Anxiety Composite	0.16 (0.11)	0.075	0.52 (0.02)
Height	0.37 (0.14)	4×10^{-4}	0.80 (0.02)
Weight	0.48 (0.14)	0.005	0.84 (0.02)

Footnote: SE – standard error; n with non-missing phenotypic data = 2806–2810 twin individuals (one co-twin per pair) for GCTA estimates and twin pairs for heritability estimates.

doi:10.1371/journal.pone.0058676.t002

Dozens of papers have been published about possible solutions to the missing heritability problem [40]. One possibility is that heritability might be overestimated in twin studies and another is that the common SNPs on commercially available DNA arrays might be missing associations due to very small effect sizes and also might be caused by rare polymorphisms of larger effect sizes [41]. Some of these issues are addressed in part by GCTA analysis. GCTA estimates overall genetic influence directly from overall SNP similarity pair by pair for a large population of unrelated individuals; in this sense, it is independent of the effect size of individual polymorphism, although it is limited to detecting the additive effects of the DNA array's common SNPs and the variants they tag. The large standard errors (Table 2) from our GCTA estimates based on a sample of 2810 indicate the daunting demands for power in trying to detect a tiny genetic signal from the noise of 1.7 million SNPs: Most of the population differ by less than 1% in overall SNP similarity across more than a million SNPs [42]. Nonetheless, for height and weight, our GCTA estimates are similar to those reported in the literature, which account for about half the heritability of these 'anchor' variables [32]. In contrast, across the five anxiety-related traits, the average GCTA estimate of 10% (Table 2) is less than one-fifth of the average twin-study heritability estimate of 55%. The total composite showed the highest, albeit non-significant, GCTA estimate but even this estimate was only about 30% of the twins study heritability estimate, which fell below the expected 50%. Importantly, consideration of the standard errors shows that if the SNPs accounted for 50% of the twin study heritability, as has been found with the 'anchor' variables height and weight, the GCTA results would have been significant in our study.

Two hypotheses for explaining the gap between these anxiety-related GCTA estimates and twin-study estimates are that GCTA underestimates genetic influence or that twin studies overestimate genetic influence, although these are not mutually exclusive hypotheses. We know that GCTA underestimates genetic influence to some extent because it only captures causal variants that are in linkage disequilibrium with the common SNPs used in the analysis; it misses the effect of rarer DNA variants not tagged by

these SNPs. In addition, GCTA only assesses additive genetic effects. So, one possibility is that anxiety is influenced by rarer DNA variants or nonadditive genetic effects to a greater extent than height and weight.

On the other hand, our twin study heritability estimates for parental ratings may be inflated – most estimates of heritability of anxiety traits in childhood and adolescence using other assessment techniques are around 30% [43], which would put our GCTA estimates more nearly in range of accounting for half the heritability. Another possibility is that, unlike in GCTA, non-additive genetic variance can inflate estimates of additive genetic variance in a twin study. That is because its estimation is generally weak without extended family data [44].

It is important to resolve this issue of the gap between GCTA and twin-study estimates of heritability in general and specifically in terms of the possibility that the gap might be larger for anxiety-related traits than for other complex traits. To the extent that GCTA estimates account for heritability, it should be possible to identify genes responsible for the heritability of anxiety using common SNPs alone if samples are sufficiently large. Larger samples could result in closing this gap by producing an increased number of significant SNP associations in GWA and by providing GCTA estimates with smaller error terms. That said, a recent study reported GCTA estimate of 0.06(0.03) for neuroticism in a sample of nearly 12,000 adults [40]. Suggesting that this gap might remain opened until either data from exome-sequencing micro-arrays are available (that tag rarer variants), or until whole-genome sequencing identifies all variants of any kind [45].

Conclusion

Our GWA results for anxiety-related traits suggest that, similar to other quantitative traits and common disorders, heritability is caused by many genes of small effect. Our GCTA results suggest that the genetic architecture of parent-rated anxiety-related traits may differ from previously published results in showing a greater gap between GCTA estimates of genetic influence and twin study estimates of heritability. One implication of knowing that there are no genes of large effect and that at least some of the genetic variance can be accounted for by the common SNPs on current DNA arrays is to increase sample sizes to detect associations of small effect size. Eventually, polygenic prediction, using composites of hundreds or thousands of DNA markers, may reach levels of predictive power useful at least for research if not for clinical practice.

Supporting Information

Text S1 Genotyping protocol, quality control and statistical analysis.
(DOC)

Table S1 Lambda Inflation rates for all variables.
(DOCX)

Acknowledgments

We gratefully acknowledge the ongoing contribution of the parents and children in the Twins Early Development Study (TEDS). We also thank the Wellcome Trust Case Consortium 2 team for their help with genotyping and quality control of our data.

Author Contributions

Conceived and designed the experiments: MT TCE RP. Performed the experiments: MT OSPD SJD KBH ELM TP. Analyzed the data: MT. Contributed reagents/materials/analysis tools: TCE OSPD CMAH TP RP. Wrote the paper: MT.

References

1. Last CG, Perrin S, Hersen M, Kazdin AE (1992) DSM-III-R anxiety disorders in children: sociodemographic and clinical characteristics. J Am Acad Child Adolesc Psychiatry 31: 1070–1076.
2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, et al. (2005) Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry 62: 593–602.
3. Ferdinand RF, Dieleman G, Ormel J, Verhulst FC (2007) Homotypic versus heterotypic continuity of anxiety symptoms in young adolescents: evidence for distinctions between DSM-IV subtypes. Journal of Abnormal Child Psychology 35: 325–333.
4. Silberg JL, Rutter M, Eaves L (2001) Genetic and environmental influences on the temporal association between earlier anxiety and later depression in girls. Biological Psychiatry 49: 1040–1049.
5. Marciniak MD, Lage MJ, Dunayevich E, Russell JM, Bowman L, et al. (2005) The cost of treating anxiety: The medical and demographic correlates that impact total medical costs. Depression and Anxiety 21: 178–184.
6. Spence SH (1997) Structure of anxiety symptoms among children: a confirmatory factor-analytic study. Journal of Abnormal Psychology 106: 280–297.
7. Gladstone GL, Parker GB, Mitchell PB, Wilhelm KA, Malhi GS (2005) Relationship between self-reported childhood behavioral inhibition and lifetime anxiety disorders in a clinical sample. Depression and Anxiety 22: 103–113.
8. Goldsmith HH, Lemery KS (2000) Linking temperamental fearfulness and anxiety symptoms: a behavior-genetic perspective. Biological Psychiatry 48: 1199–1209.
9. Franic S, Middeldorp CM, Dolan CV, Ligthart L, Boomsma DI (2010) Childhood and adolescent anxiety and depression: beyond heritability. J Am Acad Child Adolesc Psychiatry 49: 820–829.
10. Eley TC, Bolton D, O'Connor TG, Perrin S, Smith P, et al. (2003) A twin study of anxiety-related behaviours in pre-school children. The Journal of Child Psychology and Psychiatry 44: 945–960.
11. Hallett V, Ronald A, Rijsdijk F, Eley TC (2009) Phenotypic and genetic differentiation of anxiety-related behaviors in middle childhood. Depression and Anxiety 26: 316–324.
12. Kendler KS, Gardner CO, Lichtenstein P (2008) A developmental twin study of symptoms of anxiety and depression: evidence for genetic innovation and attenuation. Psychological Medicine 38: 1567–1575.
13. Silberg J, Pickles A, Rutter M, Hewitt J, Simonoff E, et al. (1999) The influence of genetic factors and life stress on depression among adolescent girls. Arch Gen Psychiatry 56: 225–232.
14. Trzaskowski M, Zavos HM, Haworth CM, Plomin R, Eley TC (2012) Stable genetic influence on anxiety-related behaviours across middle childhood. J Abnorm Child Psychol 40: 85–94.
15. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273: 1516–1517.
16. Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3: 391–397.
17. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6: 95–108.
18. WTCCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
19. Hindorff LA, Junkins HA, Mehta JP, Manolio TA (2012) A Catalog of Published Genome-Wide Association Studies. National Human Genome Research Institute.
20. Calboli FC, Tozzi F, Galwey NW, Antoniades A, Mooser V, et al. (2010) A genome-wide association study of neuroticism in a population-based sample. PloS one 5: e11504.
21. Shifman S, Bhomra A, Smiley S, Wray NR, James MR, et al. (2008) A whole genome association study of neuroticism using DNA pooling. Molecular Psychiatry 13: 302–312.
22. van den Oord EJ, Kuo PH, Hartmann AM, Webb BT, Moller HJ, et al. (2008) Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. Arch Gen Psychiatry 65: 1062–1071.
23. Maher B (2008) Personal genomes: The case of the missing heritability. Nature 456: 18–21.
24. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753.
25. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. American Journal of Human Genetics 88: 76–82.

26. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42: 565–569.
27. Oliver BR, Plomin R (2007) Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. Twin Research and Human Genetics 10: 96–105.
28. Kovas Y, Haworth CMA, Dale PS, Plomin R (2007) The genetic and environmental origins of learning abilities and disabilities in the early school years. Monographs of the Society for Research in Child Development 72: 1–144.
29. Lehmann EL (1975) Nonparametrics: statistical methods based on ranks. San Francisco: Holden-Day.
30. Marchini J, Howie BN, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies via imputation of genotypes. Nature Genetics 38: 906–913.
31. R Development Core Team (2011) R: A Language and Environment for Statistical Computing. 2.13.1 ed. Vienna, Austria.
32. Plomin R, Haworth CMA, Meaburn EL, Price T, 2WTCCC, et al. (in press) Common DNA markers can account for more than half of the genetic influence on cognitive abilities. Psychological Science.
33. Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19: 149–150.
34. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832–838.
35. Walley AJ, Asher JE, Froguel P (2009) The genetic contribution to non-syndromic human obesity. Nat Rev Genet 10: 431–442.
36. Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R (2007) Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. Mol Psychiatry 13: 729–740.
37. Docherty SJ, Kovas Y, Petrill SA, Plomin R (2010) Generalist genes analysis of DNA markers associated with mathematical ability and disability reveals shared influence across ages and abilities. BMC Genet 11: 61.
38. Davies G, Tenesa A, Payton A, Yang J, Harris SE, et al. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 16: 996–1005.
39. Davis O, Butcher L, Docherty S, Meaburn E, Curtis C, et al. (2010) A Three-Stage Genome-Wide Association Study of General Cognitive Ability: Hunting the Small Effects. Behavior Genetics 40: 759–767.
40. Vinkhuyzen AA, Pedersen NL, Yang J, Lee SH, Magnusson PK, et al. (2012) Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. Translational psychiatry 2: e102.
41. McCarthy MI, Hirschhorn JN (2008) Genome-wide association studies: potential next steps on a genetic journey. Human molecular genetics 17: R156–165.
42. Plomin R (2012) Genetics: How intelligence changes with age. Nature 482: 165–166.
43. Eley TC, Gregory A.M. (2004) Behavioral genetics. Anxiety disorders in children and adolescents. New York: Guilford Press.
44. Plomin R, DeFries JC, Knopik VS, Neiderhiser JM (2012) Behavioral Genetics. New York: Worth Publishers. 560 p.
45. Plomin R (2012) Child Development and Molecular Genetics: 14 Years Later. Child Dev 30: 1467–8624.

CHAPTER 3: NO GENETIC INFLUENCE FOR CHILDHOOD BEHAVIOUR PROBLEMS FROM DNA ANALYSIS

This chapter is presented as a published paper and is an exact copy of the following journal publication:

Maciej Trzaskowski*, Philip S. Dale and Robert Plomin (in press). *Journal of the American Academy of Child and Adolescent Psychiatry.*

* Corresponding Author: Box PO80, SGDP Centre, Institute of Psychiatry, DeCrespigny Park, London SE5 8AF

E-mail Address: maciej.trzaskowski@kcl.ac.uk

Tel: (44) 20 7848 0748

Fax: (44) 20 7848 0866

Running Title

DNA heritability of behavior problems

ABSTRACT

Objective. Twin studies of behavior problems in childhood point to substantial genetic influence. It is now possible to estimate genetic influence using DNA alone in samples of unrelated individuals, not relying on family-based designs such as twins. A linear mixed model, which incorporates DNA microarray data, has confirmed twin results by showing substantial genetic influence for diverse traits in adults. Here we present direct comparisons between twin and DNA heritability estimates for childhood behavior problems as rated by parents, teachers, and children themselves.

Method. Behavior problem data from 2500 UK-representative 12-year-old twin pairs were used in twin analyses; DNA analyses were based on one member of the twin pair with genotype data for 1.7 million DNA markers. Diverse behavior problems were assessed, including autistic, depressive, and hyperactive symptoms. Genetic influence from DNA was estimated using Genome-Wide Complex Trait Analysis (GCTA) and the twin estimates of heritability were based on standard twin model-fitting.

Results. Behavior problems in childhood – whether rated by parents, teachers or children themselves – show no significant genetic influence using GCTA, even though twin study estimates of heritability are substantial in the same sample, and even though both GCTA and twin study estimates of genetic influence are substantial for cognitive and anthropometric traits.

Conclusions. We suggest that this new type of 'missing heritability' – the gap between GCTA and twin study estimates for behavior problems in childhood – is due to nonadditive genetic influence, which will make it more difficult to identify genes responsible for heritability.

INTRODUCTION

Behavior problems in childhood – such as anxiety, depression, autistic symptoms, hyperactivity, and conduct problems – are common, with a cumulative incidence during childhood of 12% for one or more disorders[1]. They are not always transient problems that disappear as children develop: Half of all lifetime cases of diagnosed psychopathology begin in childhood[2].

Furthermore, their heritability is surprisingly high: For example, twin studies using parent ratings typically report heritabilities in the range of 40% for anxiety and depression to 60% for autistic symptoms and hyperactivity[3]. Consequently, childhood behavior problems have become the target of genome-wide association (GWA) studies that attempt to identify the genes responsible for their heritability. As in other life sciences[4], these GWA expeditions have come up largely empty-handed[5]. This 'missing heritability' is the key puzzle in DNA research on complex traits and common disorders – only a small portion of genes responsible for their heritability has been identified[5,6]. Although attention has focused on the difficulties in identifying the many genes of small effect responsible for heritability, the other side of the missing heritability gap is the heritability estimate itself, which has relied on family-based studies of twins and adoptees[3].

It is now possible to test the validity of these heritability estimates using DNA from unrelated individuals, a method called *Genome-wide Complex Trait Analysis*[7,8]. GCTA research has shown that the common DNA variants (single-nucleotide polymorphisms, SNPs) genotyped on DNA arrays used in GWA studies yield substantial estimates of genetic influence for height[7] and weight[9], psychiatric and medical disorders[10-12], personality[13], and cognitive traits[14-16]. However, these GCTA estimates of genetic influence cannot completely close the heritability gap, in part because GCTA is limited to additive effects of causal variants tagged by the common SNPs on current DNA arrays used in GWA research[8].

For the first time, we report GCTA estimates for childhood behavior problems as rated by parents, teachers and the children themselves. In order to gauge the true breadth of the missing heritability gap, we

compared GCTA estimates to twin heritability estimates for the same measures in the same sample. We also compared these results for behavior problems to results for height and weight and cognitive traits in the same sample.

METHOD

SAMPLE AND GENOTYPING

The sample was drawn from the Twins Early Development Study (TEDS), which is a multivariate longitudinal study that recruited over 11,000 twin pairs born in England and Wales in 1994, 1995 and 1996[17]. TEDS has been shown to be representative of the UK population[18]. The project received approval from the Institute of Psychiatry ethics committee (05/Q0706/228) and parental consent was obtained prior to data collection.

The present analyses were limited to children for whom DNA, genome-wide genotyping, and behavior problem and cognitive data were available. Moreover, the twin analyses were based only on twins included in the GCTA analyses in order to provide a more precise comparison between GCTA and twin study results.

DNA was available for 3747 11- and 12-year-old children (11.5 average age) whose first language was English and had no major medical or psychiatric problems. From that sample, DNA samples of 3665 individuals (only one member of a twin pair) were successfully hybridized to Affymetrix GeneChip 6.0 SNP genotyping arrays using standard experimental protocols as part of the WTCCC2 project. In addition to nearly 700,000 genotyped SNPs, more than one million other SNPs were imputed from HapMap 2, HapMap 3 and WTCCC controls, using IMPUTE v.2 software[19]. 3152 DNA samples (1446 males and 1706 females) survived quality control criteria for ancestry, heterozygosity, relatedness, and hybridization intensity outliers. To control for ancestral stratification, we performed principal component analyses on a subset of 100,000 quality-controlled SNPs after removing SNPs in linkage disequilibrium ($r^2 > 0.2$)[20]. Using the Tracy-Widom Test[21], we identified 8 axes with $p < 0.05$, which were used as covariates in GCTA analyses.

Of these 3152 children, the present analyses were limited to those for whom behavior problem and cognitive data were available. Twin zygosity was diagnosed on the basis of physical similarity and questionable cases were verified with analysis of DNA markers[18]. As expected, approximately equal numbers of MZ, DZ same-sex, and DZ opposite-sex twins were included; DZ same-sex and opposite-sex pairs were combined in order to increase power and because previous twin analyses of these data show no evidence of qualitative or quantitative sex differences in sex-limitation models[22]. For the measures of behavior problems, the numbers of individuals for GCTA analyses range from 2687 to 2698 for self-report, 2687 to 2700 for parent ratings, and 2034 to 2139 for teacher ratings. The numbers of pairs of twins range from 2668 to 2683 for self-report, 2680 to 2695 for parent ratings, and 1783 to 1925 for teacher ratings. The sample sizes for the GCTA results shown are 2325 for 'g' and language, 2238 for 'g' and mathematics, 2250 for 'g' and reading, and 2296 for height and weight.

Measures

All of the measures have been reported in previous TEDS publications which can be consulted for greater detail[23].

Behavior problems

The behavior problem measures described in this section have been widely used in the literature. As is the case in the literature, these measures are modestly correlated, 0.33 on average for the scores described below. Given the modest correlation among the measures and the focus of the present paper on comparisons between GCTA and twin estimates for diverse behavior problems, we present results separately for the behavior problem scales rather than conducting multivariate analyses.

Conners (ADHD). ADHD symptoms were assessed via parent-rated questionnaire, which was the DSM-IV-based ADHD scale from the Conners' Parent Rating Scale-Revised (CPRS-R). The questionnaire consisted of two scales: inattentiveness and hyperactivity-impulsivity.

APSD (psychopathic symptoms). Antisocial behavior was assessed using parent and teacher ratings on the Antisocial Process Screening Device. The questionnaire included three scales: Callous-Unemotional, Impulsivity and Narcissism.

CAST (autistic-like symptoms). The Childhood Asperger Syndrome Test questionnaire was rated by parents and teachers and includes three scales: Communication, Non-social and Social, from which a composite was also formed.

MFQ (depressive symptoms). The Moods and Feelings Questionnaire was rated by the children and their parents.

SDQ (behavior problems). The strengths and Difficulties Questionnaire was assessed by the children and their parents. The questionnaire includes four behavior problem scales (anxiety, conduct, hyperactivity, peer problems) from which a composite was created. The SDQ also includes a positive prosocial scale that was not included in these analyses of behavior problems.

Cognitive tests

Cognitive data were collected online via the Internet using adaptive branching, which enabled measurement of the full range of ability using a relatively small number of items.

Reading. Four measures of reading were employed. Two measures assessed reading comprehension: the reading comprehension subtest of the Peabody Individual Achievement Test (PIAT) and the GOAL Formative Assessment in Literacy for Key Stage 3. Reading fluency was assessed by an adaptation of the

Woodcock-Johnson III Reading Fluency Test (WJRF) and by the Test of Word Reading Efficiency (TOWRE), which was administered by telephone.

Mathematics. Assessment of mathematics targeted three components of mathematics: Understanding Number, Non-numerical Processes, and Computation and Knowledge. The items for these three scales were based on the National Foundation of Educational Research 5-14 Mathematics Series

Language. Three components of language were assessed: syntax, semantics and pragmatics. Syntax was measured using the Listening Grammar subtest of the Test of Adolescent and Adult Language. Semantics was assessed using Level 2 of the Figurative Language subtest of the Test of Language Competence. Pragmatics was assessed using Level 2 of the Making Inferences subtest of the Test of Language Competence.

Verbal, non-verbal and general cognitive abilities. The verbal tests were WISC-III-PI Multiple Choice Information (General Knowledge) and Vocabulary Multiple Choice subtest. The two non-verbal reasoning tests were WISC-III-UK Picture Completion and Raven's Standard and Advanced Progressive Matrices. General cognitive ability was indexed as a composite of the four verbal and non-verbal tests.

Height and weight

Height and weight were assessed at age 12 via self-report.

Composite measures

To create composite scores for the measures, standardized residuals were derived for each scale regressed on sex and age. Outliers above or below 3 SD from the mean were excluded and the scale was quantile normalized[24,25]. The composites were created as unit-weighted means requiring complete data for more than half of the measure's scales (i.e., 3 out of 4 or 2 out of 3 scales). All procedures were executed using R (www.r-project.org[26]).

STATISTICAL ANALYSES

Genome-wide Complex Trait Analysis (GCTA). GCTA software was used to conduct these analyses[8]. Of note, although GCTA is a software package, for simplicity we will refer to the full process of estimating genetic influence from SNP data simply as GCTA. Before the variance of a trait can be decomposed, the first step is to calculate pairwise genomic similarity between all pairs of individuals in the sample using all genetic markers genotyped or imputed from the SNP array. Because GCTA is designed to estimate genetic variance due to close linkage disequilibrium between unknown causal variants and genotyped SNPs from a sample of unrelated individuals in the population, any close genetic relatedness is eliminated; for this reason any individual whose genetic similarity is equal to or greater than a third or fourth cousin is removed (estimate of pairwise relatedness > 0.025).

Conceptually, when performing GCTA analysis, we compare a matrix of pairwise genomic similarity to a matrix of pairwise phenotypic similarity using a random-effects mixed linear model[7]. In univariate analysis, the variance of a trait can be partitioned using residual maximum likelihood into genetic and residual components. Therefore the residual component includes any source of variance that is not an additive effect of common SNPs, including non-additive genetic effects, rare variants, environment, gene-environment interaction and error. Detailed description of this method can be found in Yang et al[7,8]. The eight principal components described earlier were used as covariates; as mentioned in the previous section, all phenotypes were age- and sex-regressed prior to analysis.

Twin analysis. In contrast to GCTA, the twin method models variance/covariance for pairs of related individuals: monozygotic (MZ) twins who are genetically 100% identical, and dizygotic (DZ) twins who share on average 50% of their segregating alleles. Differences in within-pair correlations for MZ and DZ twins are then used to partition the variance into genetic and environmental effects. Variance is attributed to additive genetic influence to the extent that MZ correlations are higher than DZ twins. The twin method, unlike GCTA, partitions environmental influence into two components -- shared or common (C) environmental influences, which is residual MZ twin resemblance not explained by genetics, and non-

shared or unique environmental influences (E), which is the extent to which MZ twins differ and includes error of measurement. Detailed description of this ACE model and the discussion of related issues can be found elsewhere[3].

The twin data were modeled using Cholesky decomposition of the variance within structural equation modeling software OpenMx[27]. Standard univariate model-fitting procedures were followed, as described in previous TEDS publications[18]; standard errors were derived from 95% confidence intervals. Although in most cases the influence of shared environment was not significant, the full ACE model was used for the comparison with the GCTA results. Although C was not significantly different from zero, dropping it from the model could have inflated A, which would confound the comparison between twin and GCTA results. It should be noted that GCTA does not discriminate C and E; both C and E are included in the GCTA estimate of non-genetic residuals. In other words, the 'A' and 'E' of GCTA and the 'A' and 'E' of twin ACE model-fitting are not the same. For GCTA, 'A' denotes additive effects of DNA variants tagged by the common SNPs on our DNA array, and 'E' includes all residual variance. In contrast, in twin analysis, 'A' includes additive genetic effects of any DNA sequence differences, not just common SNPs; variance not explained by A is partitioned into C and E.

RESULTS

Figure 1 compares GCTA and twin study estimates for the anchor variables of height and weight as well as for cognitive traits. These results are based on the same individuals and twin pairs used in the present analyses of behavior problems, although the results are highly similar to those previously published for the entire TEDS sample[16]. As expected from the literature, the twin study heritability estimates for height and weight are about 80% and the estimates for the cognitive traits are about 50% (~40% to 60%). The GCTA estimates are about 40% for height and weight and about 25% (~20% to 30%) for the cognitive traits. All of the GCTA estimates are statistically significant, as indicated by the standard errors. These significant and substantial GCTA estimates have two important implications. First, they validate the twin method. Second, they imply that sufficiently large GWA studies using current DNA arrays limited to additive effects of

common SNPs should be able to account for about 50% of the heritability for height, weight, and cognitive traits. The finding that GCTA estimates are only half the twin heritability estimates is similar to previous reports for these variables and could be due to several factors that either result in GCTA underestimates of twin heritability – such as non-additive gene-gene interactions, gene-environment interactions, rare alleles – or factors that lead to inflation of heritability estimates in twin studies[5].

Figure 2 compares GCTA and twin study estimates of heritability for composite measures of behavioral problems for self-report, parent ratings, and teacher ratings. Results for the scales that comprise these composites as well as the full variance decomposition are included in the Online-Only Supplement (Table S1). Twin heritability estimates are similar to those reported in the literature: About 40% heritability for self-report and about 60% heritability for parent and teacher ratings. In contrast, GCTA estimates are non-significant and mostly zero for self-report and parent measures of behavior problems. For teacher ratings, a hint of genetic influence emerged, although these GCTA estimates of about 10% are not nearly statistically significant, as indicated by the standard errors. The standard errors are larger for GCTA estimates than for twin estimates because GCTA is based on slight (less than 2.5%) overall pair-by-pair differences in genetic similarity across the 1.7 million SNPs genotyped from the DNA array, whereas the twin estimate is based on the comparison of 100% genetic similarity for MZ twins and 50% similarity for DZ twins for additive genetic effects. However, if the GCTA estimates for behavior problems were half the twin estimates of heritability, as in the case of height and weight and cognitive traits (Figure 1), the GCTA analysis would have adequate power to detect them, as indicated by the standard errors.

DISCUSSION

Why do GCTA estimates show no significant genetic influence for diverse childhood behavior problems as rated by parents, teachers or children themselves, even though twin study estimates of heritability are significant and substantial in the same sample using the same measures, and even though GCTA estimates for cognitive traits are significant and substantial? One broad category of explanations involves mechanisms by which GCTA underestimates twin heritability, more so for behavior problems than for

cognitive traits. As mentioned earlier, GCTA underestimates twin heritability because it only captures additive genetic effects tagged by the common SNPs used on GWA arrays. Gene-gene interactions, gene-environment interactions, and rare alleles will widen the gap between GCTA and twin estimates of heritability. However, it is not clear why this gap would be greater for behavior problems than for cognitive traits.

Nonetheless, we can test one of these hypotheses -- that nonadditive genetic variance is greater for behavior problems than for cognitive traits -- in our study. Because MZ twins are identical genetically, they are identical even for interactions among many genes (epistasis), whereas such epistatic effects, on average, scarcely contribute to similarity for DZ twins[3]. Additive genetic effects contribute to MZ twins being twice as similar as DZ twins, whereas the hallmark of nonadditive genetic variance is that MZ twins are more than twice as similar as DZ twins. However, in our study there is no evidence for nonadditive genetic influence either for behavior problems or for cognitive traits: In all cases, MZ correlations are no more than twice as similar as DZ twins. For example, for total behavior problems, MZ and DZ correlations are 0.56 and 0.32, respectively, for self-reports in adolescence, 0.80 and 0.52 for parent ratings, and 0.61 and 0.31 for teacher ratings. Nonetheless, as discussed later, it is possible that nonadditive genetic effects for behavior problems are masked by other factors.

Another, non-mutually exclusive, explanation of why GCTA underestimates twin heritability for behavior problems more than for cognitive abilities is that additive genetic variance might be greater for cognitive traits than for behavior problems. Not previously considered in this context is assortative mating, which increases additive genetic variance for all loci associated with the trait for which spouses correlate. Assortative mating, indexed by the correlation between spouses, is often greater for cognitive traits than for behavior problems: Spouse correlations are about 0.40 for cognitive ability[28]. In contrast, assortative mating was reported 0.00 for autistic symptoms[29] and 0.02 for hyperactivity in one study[29], 0.38 and 0.11 respectively in other studies[30,31]. Because assortative mating is trait-specific and increases additive genetic variance cumulatively over generations, the greater assortative mating for cognitive traits than for behavior problems would increase additive genetic variance for cognitive traits but not for behavior problems.

Because GCTA only detects additive genetic variance, assortative mating could thus account for the higher GCTA estimates for cognitive traits than for behavior problems. However, an obstacle for this hypothesis is that more additive genetic variance for cognitive traits than for behavior problems should lead to higher heritabilities for cognitive traits, but the results in Figures 1 and 2 show that this is not the case, although this effect could also be masked by countervailing effects, as discussed later.

A second broad category of explanations, and again not mutually exclusive, is that twin studies overestimate heritability for behavior problems more than for cognitive traits. One reason to take this seriously is that twin studies yield higher estimates of heritability than do adoption studies for personality traits, which are related to behavior problems in that personality includes traits such as emotionality, impulsivity, and activity level[32]. Moreover, the first report of GCTA estimates for personality supported the adoption results with estimates of about 10%, the lowest reported GCTA estimates for any domain of behavior prior to the present study of behavior problems[13]. In contrast, heritability estimates are similar for twin and adoption studies of cognitive traits[5]. It has been suggested that the twin/adoption heritability difference for personality is due to greater nonadditive genetic variance for personality than for cognitive traits because estimates of heritability from twin studies include nonadditive as well as additive genetic variance (broad heritability), whereas adoption studies that involve first-degree relatives are largely limited to additive genetic variance (narrow heritability)[32]. If nonadditive genetic variance is the solution to the twin/adoption heritability difference for personality, it would imply that designs that do not involve MZ twins underestimate heritability caused by nonadditive genetic effects. In other words, twin studies do not overestimate heritability as compared to adoption designs; instead, adoption designs involving first-degree relatives estimate narrow heritability while twin studies estimate broad heritability. Although this means that nonadditive genetic variance again emerges as a good candidate for explaining the GCTA/twin heritability gap, this explanation goes against the pattern of MZ and DZ twin correlations for behavior problems in childhood in the present study which shows no evidence for nonadditive genetic variance, as indicated earlier.

Another methodological possible explanation of the low GCTA heritability estimates could be the skewed distributions, which are often found for measures of behavior problems. As shown in Supplement Figure S1, some of the untransformed distributions are skewed but the transformed distributions, which were used in our analyses, are normal. In order to check on the possibility that the transformation could affect GCTA and twin estimates to different extents, we compared results for both methods using transformed and untransformed scales and found little difference (see footnote in Figure S1).

Overall, greater nonadditive genetic influence for behavior problems than for cognitive traits emerges as the leading candidate to explain the greater GCTA/twin heritability gap for behavior problems. The only problem with this explanation is that our twin results for behavior problems in children do not indicate nonadditive genetic effects, even though other twin studies of behavior problems in children and twin studies of adult personality point to some nonadditive genetic effects. In the absence of a more parsimonious explanation, we suggest that nonadditive genetic effects for behavior problems in childhood are masked by a general inflation of twin similarity for both MZ and DZ twins. One possibility is that this general twin inflation could be due to experiences that are shared by members of both MZ and DZ twin pairs. However, this possibility seems less unlikely because such shared experiences would seem likely to affect cognitive traits at least as much as behavior problems. A general inflation of twin similarity due to rating bias is another, more promising, possibility: A major difference between behavior problems and cognitive traits is that behavior problems are rated on questionnaires whereas cognitive traits are measured by tests and ratings are inherently more prone to bias than tests. Such a general inflation of MZ and DZ twin correlations would mask nonadditive genetic variance because it would reduce the difference between MZ and DZ twin correlations. For example, suppose that the 'true' MZ and DZ twin correlations were 0.3 and 0.1, respectively, suggesting some nonadditive genetic variance. Inflating both twin correlations by 0.1 would result in MZ and DZ correlations of 0.4 and 0.2, respectively, suggesting only additive genetic influence.

Figure 3 illustrates our hypothesis in the context of missing heritability. It introduces a second type of missing heritability. The familiar missing heritability is the extent to which the cumulative effect of all SNPs

identified in GWA studies fall short of accounting for twin study heritability estimates. This could be called *missing GWA heritability* to distinguish it from *missing GCTA heritability*, which is the extent to which GCTA estimates fall short of accounting for twin study heritability estimates, and which sets the limit for GWA heritability because both GCTA and GWA are limited to detecting additive effects of common SNPs. We propose that twin studies do not overestimate heritability for behavior problems. Rather, twin studies accurately detect nonadditive as well as additive genetic variance for behavior problems, but mostly additive genetic variance for cognitive traits. We do not suggest that all genetic variance for behavior problems is nonadditive, only that there is relatively more nonadditive genetic variance for behavior problems (masked by inflation of twin correlations for both MZ and DZ twins in the present study) and relatively more additive genetic variance for cognitive traits. If the heritability of behavior problems is about 50% and about half of the heritability is due to nonadditive genetic variance, whereas most of the genetic variance for cognitive traits is additive, this would explain the present results, as illustrated in Figure 3. If GCTA heritability for behavior problems were about half the GCTA heritability for cognitive traits (e.g., 12% vs 25% respectively in Figure 3), the present study would have little power to detect it, as indicated by the standard errors in Figure 2.

In summary, we propose that a combination of three factors are responsible for the greater GCTA/twin heritability gap for behavior problems as compared to cognitive traits. First, nonadditive genetic variance is greater for behavior problems than for cognitive traits, reducing GCTA estimates for behavior problems relative to cognitive traits. Second, greater assortative mating for cognitive traits as compared to behavior problems produces more additive genetic variance for cognitive traits, which results in greater GCTA estimates and thus lowers the GCTA/twin heritability gap for cognitive traits but not for behavior problems. Third, the reason our twin data do not indicate nonadditive genetic effects for behavior problems in childhood is that these nonadditive genetic effects are masked by inflated correlations for both MZ and DZ twins for ratings of behavior problems. Although this set of hypotheses is speculative, if true, it suggests that GWA studies will find it more difficult to identify SNPs associated with behavior problems than with cognitive traits. It also suggests that nonadditive genetic variance might contribute importantly to the

genetic architecture of behavior problems in childhood, and perhaps in adult personality and psychopathology as well.

ACKNOWLEDGEMENTS

Twins Early Development Study (TEDS) is supported by a program grant to RP from the UK Medical Research Council (G0901245, and previously G0500079), with additional support from the US National Institutes of Health (HD044454; HD046167). Genome-wide genotyping was made possible by a grant from the Wellcome Trust to the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z; 085475/Z/08/Z). RP is supported by a Medical Research Council Research Professorship award (G19/2) and a European Advanced Investigator award (295366); MT is supported by a Medical Research Council studentship. First author had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

1. Merikangas KR, He J-P, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics.* 2010;125(1):75-81.

2. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62(6):593-602.

3. Plomin R, DeFries JC, Knopik VS, Neiderhiser JM. *Behavioral genetics.* 6th ed. New York, NY: Worth Publishers; 2012.

4. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet.* 2012;90(1):7-24.

5. Plomin R. Child development and molecular genetics: 14 years later. *Child Dev.* 2012. doi: 10.1111/j.1467-8624.2012.01757.x

6. Maher B. Personal genomes: The case of the missing heritability. *Nature.* 2008;456(7218):18-21.

7. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42(7):565-569.

8. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet.* 2011;88(1):76-82.

9. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA,

Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME, Visscher PM. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet.* 2011;43(6):519-525.

10. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet.* 2011;88(3):294-305.

11. Lee SH, DeCandia TR, Ripke S, Yang J, Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ), International Schizophrenia Consortium (ISC), Molecular Genetics of Schizophrenia Collaboration (MGS), Sullivan PF, Goddard ME, Keller MC, Visscher PM, Wray NR. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet.* 2012;44(3):247-250.

12. Lubke GH, Hottenga JJ, Walters R, Laurin C, de Geus EJ, Willemsen G, Smit JH, Middeldorp CM, Penninx BW, Vink JM, Boomsma DI. Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol Psychiatry.* 2012;72:707-709.

13. Vinkhuyzen AAE, Pedersen NL, Yang J, Lee SH, Magnusson PK, Iacono WG, McGue M, Madden PA, Heath AC, Luciano M, Payton A, Horan M, Ollier W, Pendleton N, Deary IJ, Montgomery GW, Martin NG, Visscher PM, Wray NR. Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Transl Psychiatry.* 2012;2:e102.

14. Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, Ke X, Le Hellard S, Christoforou A, Luciano M, McGhee K, Lopez L, Gow AJ, Corley J, Redmond P, Fox HC, Haggarty P, Whalley LJ, McNeill G, Goddard ME, Espeseth T, Lundervold AJ, Reinvang I, Pickles A, Steen VM, Ollier W, Porteous DJ, Horan M, Starr JM, Pendleton N, Visscher PM, Deary IJ. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry.* 2011;16(10):996-1005.

15. Deary IJ, Yang J, Davies G, Harris SE, Tenesa A, Liewald D, Luciano M, Lopez LM, Gow AJ, Corley J, Redmond P, Fox HC, Rowe SJ, Haggarty P, McNeill G, Goddard ME, Porteous DJ, Whalley LJ, Starr JM,

Visscher PM. Genetic contributions to stability and change in intelligence from childhood to old age. *Nature.* 2012;482(7384):212-215.

16. Plomin R, Haworth CMA, Meaburn EL, Price T, 2 WTCCC, Davis OSP. Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychol Sci.* in press.

17. Oliver BR, Plomin R. Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems from childhood through adolescence. *Twin Res Hum Genet.* 2007;10(1):96-105.

18. Kovas Y, Haworth CMA, Dale PS, Plomin R. The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monogr Soc Res Child Dev.* 2007;72:1-144.

19. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.

20. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB. A whole-genome association study of major determinants for host control of HIV-1. *Science.* 2007;317(5840):944-947.

21. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):2074-2093.

22. Greven C, Rijsdijk F, Plomin R. A twin study of ADHD symptoms in early adolescence: Hyperactivity-impulsivity and inattentiveness show substantial genetic overlap but also genetic specificity. *J Abnorm Child Psychol.* 2011;39(2):265-275.

23. Haworth CMA, Davis OSP, Plomin R. Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioural development from childhood to young adulthood. *Twin Res Hum Genet.* 2012. doi:10.1017/thg.2012.91

24. van der Waerden BL. On the sources of my book "Moderne Algebra". *Historia Mathematica.* 1975;2:31-40.

25. Lehmann EL. *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day; 1975.

26. R Core Development Team. R: A language and environment for statistical computing (Version 2.13.1). 2011; http://www.R-project.org/. Accessed November 3, 2012.

27. Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J. OpenMx: An open source extended structural equation modeling framework. *Psychometrika.* 2011;76(2):306-317.

28. Jensen AR. Genetic and behavioural effects of nonrandom mating. In: Osbourne RT, Noble CE, Weyl N, eds. *Human variation: The biopsychology of age, race, and sex*. New York: Academic Press; 1978:51-105.

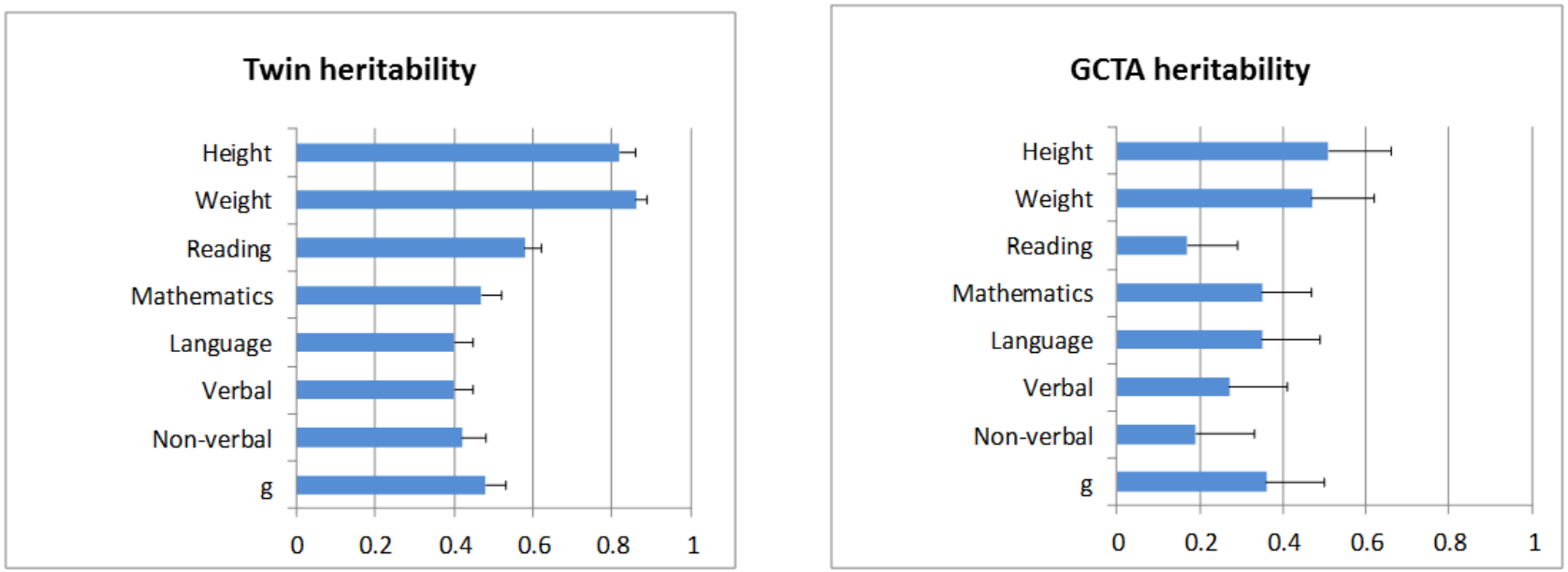
29. van Steijn DJ, Richards JS, Oerlemans AM, de Ruiter SW, van Aken MA, Franke B, Buitelaar JK, Rommelse NN. The co-occurrence of autism spectrum disorder and attention-deficit/hyperactivity disorder symptoms in parents of children with ASD or ASD with ADHD. *J Child Psychol Psychiatry.* 2012;53(9):954-963.

30. Constantino JN, Todd RD. Intergenerational transmission of subthreshold autistic traits in the general population. *Biological Psychiatry*. 2005;57(6):655-660.

31. Boomsma DI, Saviouk V, Hottenga J-J, Distel MA, de Moor MH, Vink JM, Geels LM, van Beek JH, Bartels M, de Geus EJ, Willemsen G. Genetic epidemiology of Attention Deficit Hyperactivity Disorder (ADHD Index) in adults. *PLoS One.* 2010;5:e10621.

32. Plomin R, Corley R, Caspi A, Fulker DW, DeFries JC. Adoption results for self-reported personality: Evidence for nonadditive genetic effects? *J Pers Soc Psychol.* 1998;75:211-218.

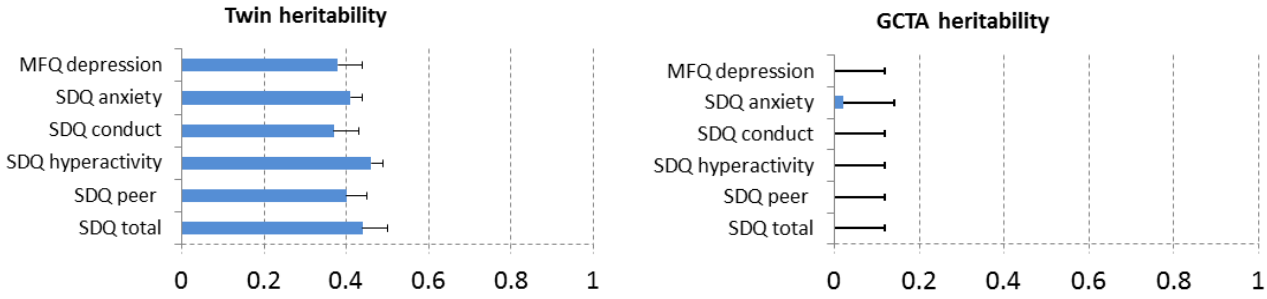
Figure 1. Genetic estimates for height, weight and cognitive trait composites from twin analyses and from GCTA. 'g' refers to general cognitive ability which is a composite of verbal and non-verbal ability.



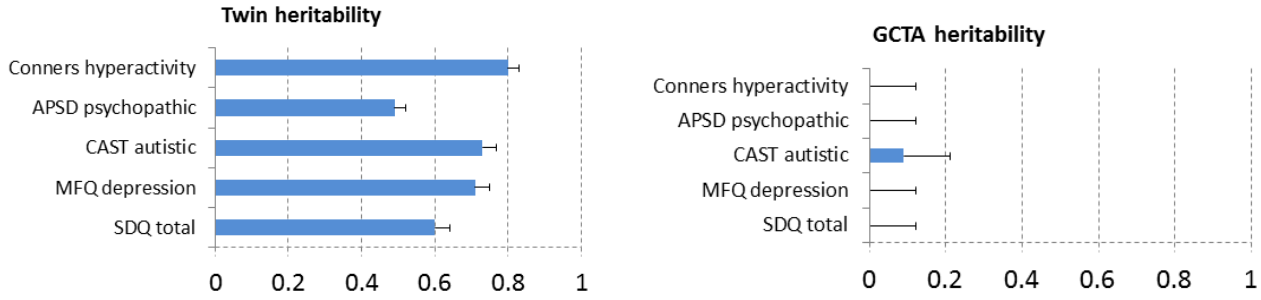
N = 2153 to 2659 twin pairs for twin analyses and N = 2281 to 2809 unrelated individuals for GCTA. Error bars in the figure indicate Standard Errors.

Figure 2. Genetic estimates for composite measures of behavior problems from twin analyses and from GCTA

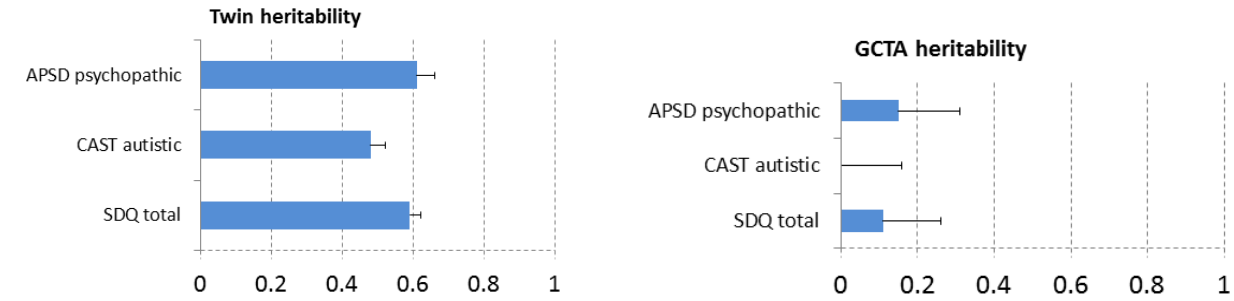
(A) Self-report



(B) Parent ratings

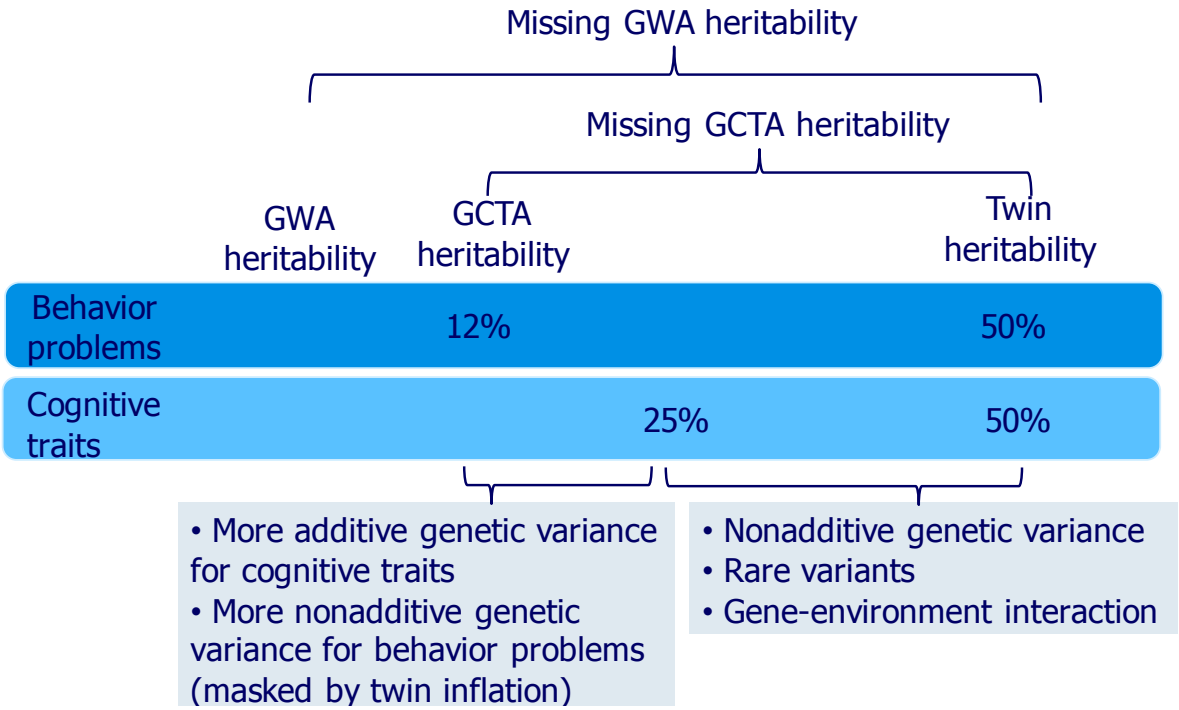


(C) Teacher ratings



Footnote: (A) self-report (N = 2153 to 2659 twin pairs for twin analyses; N = 2281 to 2809 unrelated individuals for GCTA. (B) parent ratings (N = 2680 to 2695 twin pairs for twin estimates; N = 2687 to 2700 individuals for GCTA estimates. (C) teacher ratings (N = 1783 to 1925 twin pairs for twin analyses; N = 2034 to 2139 individuals for GCTA estimates. Error bars in the figure indicate Standard Errors. Results for the constituent scales for these composites are presented in the Online-Only Supplement (Table S1).

Figure 3. Missing GWA heritability and missing GCTA heritability for behavior problems and cognitive traits



CHAPTER 4: FINDING THE MISSING HERITABILITY IN PAEDIATRIC OBESITY: THE CONTRIBUTION OF GENOME-WIDE COMPLEX TRAIT ANALYSIS

This chapter is presented as a published paper and is an exact copy of the following journal publication:

Clare H Llewellyn*, Maciej Trzaskowski*, Robert Plomin and Jane Wardle (2013). *International Journal of Obesity.*

* These authors contributed equally

OPEN

International Journal of Obesity (2013), 1–4

www.nature.com/ijo

SHORT COMMUNICATION

Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis

CH Llewellyn[1,2,3], M Trzaskowski[2,3], R Plomin[2] and J Wardle[1]

Known single-nucleotide polymorphisms (SNPs) explain $<2\%$ of the variation in body mass index (BMI) despite the evidence of $>50\%$ heritability from twin and family studies, a phenomenon termed 'missing heritability'. Using DNA alone for unrelated individuals, a novel method (in a software package called Genome-wide Complex Trait Analysis, GCTA) estimates the total additive genetic influence due to common SNPs on whole-genome arrays. GCTA has made major inroads into explaining the 'missing heritability' of BMI in adults. This study provides the first GCTA estimate of genetic influence on adiposity in children. Participants were from the Twins Early Development Study (TEDS), a British twin birth cohort. BMI s.d. scores (BMI-SDS) were obtained from validated parent-reported anthropometric measures when children were about 10 years old (mean = 9.9; s.d. = 0.84). Selecting one child per family ($n = 2269$), GCTA results from 1.7 million DNA markers were used to quantify the additive genetic influence of common SNPs. For direct comparison, a standard twin analysis in the same families estimated the additive genetic influence as 82% (95% CI: 0.74–0.88, $P<0.001$). GCTA explained 30% of the variance in BMI-SDS (95% CI: 0.02–0.59; $P = 0.02$). These results indicate that 37% of the twin-estimated heritability (30/82%) can be explained by additive effects of multiple common SNPs, and provide compelling evidence for strong genetic influence on adiposity in childhood.

International Journal of Obesity advance online publication, 26 March 2013; doi:10.1038/ijo.2013.30

Keywords: Genome-wide Complex Trait Analysis (GCTA); missing heritability; children; twins; genetics

INTRODUCTION

Family history has long been recognised as an important risk factor for obesity.[1,2] Quantitative genetic analyses using twin, family and adoption designs have demonstrated that familial resemblance in body mass index (BMI) is largely due to genetic similarity, with high heritability estimates reported from twin studies (47–90%), and moderate-to-high estimates from family (24–81%) and adoption studies (20–60%).[3,4] A recent meta-analysis of twin studies showed that heritability estimates were on average 0.07 higher in children than in adults.[3]

Genome-wide Association Studies (GWAS) have made significant headway in identifying single-nucleotide polymorphisms (SNPs) that are related to the relative body weight, indexed using BMI.[5] A large meta-analysis of 123 865 adults from 46 studies with follow-up in another 125,931 participants conducted by the Genetic Investigation of Anthropometric Traits (GIANT) consortium identified 32 SNPs robustly associated with adult BMI.[5] The majority of these SNPs (23–28) demonstrated directionally consistent effects in age- and sex-adjusted BMI in children and adolescents.[5,6] A subsequent meta-analysis of 14 studies with 5530 cases of obesity and 8318 controls identified another two SNPs associated with childhood and adolescent obesity that also showed directionally consistent effects in the previous meta-analysis of adult BMI.[5,7]

However, even in combination, the 32 established SNPs explain $<2\%$ of the variation in BMI in either adults or children,[5] although there are some suggestions that the size of the association between combined genetic obesity risk and adiposity may vary over the lifespan, peaking during late childhood (age 11) and early adulthood (age 20)[8] in line with heritability estimates. The mismatch between the high heritability estimates from quantitative genetic analyses and the small proportion of variation explained through GWAS findings across many complex traits have come to be known as the problem of 'missing heritability'.[9] Part of the missing heritability is likely to be due to rare genetic variants and some non-additive genetic effects. These contribute to the estimated genetic effect in quantitative genetic studies, but are not detected in GWAS analyses that only capture additive effects of common SNPs with minor allele frequencies of $\geqslant 5\%$. A second possibility is that there are multiple additional common genetic variants that contribute to the genetic effect observed in quantitative genetic studies, but have such small effect sizes that they cannot be detected even in the huge data sets used in contemporary GWAS analyses. However, until there is direct molecular genetic evidence for these additional sources of genetic influence, missing heritability is not clarified, and questions will remain about whether the heritability of obesity has been overestimated by quantitative genetic studies.

A novel approach called Genome-wide Complex Trait Analysis (GCTA) takes advantage of the fact that the degree of genetic resemblance for common SNPs at the whole-genome level is normally distributed among unrelated individuals. This can be used to quantify the proportion of the variation in a particular phenotype that is explained by the total common SNP similarity, effectively a molecular genetic estimate of heritability.[10] The purpose of GCTA is not to identify specific SNPs related to the target phenotype, but rather to estimate the total additive genetic effect of the common SNPs used on currently available DNA

[1]Department of Epidemiology and Public Health, Health Behaviour Research Centre, University College London, London, UK and [2]MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK. Correspondence: Dr CH Llewellyn, Department of Epidemiology and Public Health, Health Behaviour Research Centre, University College London, Gower Street, London WC1E 6BT, UK.
E-mail: c.llewellyn@ucl.ac.uk
[3]Authors contributed equally to this work
Received 22 November 2012; revised 24 January 2013; accepted 28 January 2013

arrays. Its value relative to GWAS comes from the fact that the GCTA estimate includes the effect of SNPs well below the current GWAS threshold.

GCTA has made major inroads into explaining the missing heritability of adiposity in adults. The first report found a genetic effect due to additive effects of common SNPs of 16.5%;[11] a remarkable order-of-magnitude increase compared with the effect of known genetic variants, and not far off the lower limit for additive genetic influence as estimated from family studies (e.g.[12]). A second study produced very similar results, with GCTA estimates of 14 and 10% in two independent adult samples.[13] In this study, we provide the first pediatric GCTA estimate of additive genetic effects on adiposity in a sample of unrelated children. We also include the twin-based estimate of heritability in the same sample for direct comparison with the GCTA estimate by including data from the co-twin in the same families.

SUBJECTS AND METHODS

Sample

Data were from the Twins Early Development Study (TEDS), a population-based cohort of monozygotic and dizygotic twins (>11 000 pairs) born between 1994 and 1996 in England and Wales.[14] Twins and parents provided informed consent for each part of the study prior to data collection. King's College London's Ethics Committee provided ethical approval.

Genotyping

Genome-wide genotyping was completed in 2010 for one randomly selected child in each of 3665 families; of these, 3152 (1446 male and 1706 female subjects) survived quality control criteria for ancestry, heterozygosity, relatedness and hybridisation intensity outliers (for details see[15]). Genotyping and quality control was done using the Affymetrix 6.0 GeneChip SNP genotyping array (Affymetrix Inc, Santa Clara, CA, USA) using standard experimental protocols as part of the WTCCC2 project.[16] SNPs were selected on their minor allele frequency (>0.01), genotype call-rate (>0.80), Hardy–Weinberg Equilibrium ($>10^{-20}$) and plate effect *P*-value ($>10^{-6}$), which resulted in ~700 000 quality-controlled genotyped SNPs. In addition, there were ~2.5 million SNPs imputed from HapMap 2 and 3, and WTCCC controls, using the programme IMPUTE v.2 software.[17] Imputed SNPs were screened using much more stringent quality control that resulted in reduction to ~1 000 000 SNPs, giving a total of 1.7 million (quality controled) SNPs (for details see[15]). To control for ancestral stratification, we performed principal component analysis using EIGENSTRAT from EIGENSOFT package[18] and identified significant axes using the Tracy–Widom Test.[19] This resulted in eight axes with $P<0.05$ that were used as covariates in GCTA analyses.

Measurement of adiposity

Height and weight data were obtained in 2005 when the children were 8–11 years old, as part of a study of the heritability of adiposity.[20] Parents were sent detailed instructions and asked to record each child's weight to the nearest pound or tenth of a kg, and height to the nearest cm, as well as the date of measurement. Parent- and researcher-measured heights and weights were correlated 0.90 and 0.83 in a subsample of 228 families.[20] BMI was calculated as weight (kg)/height (m)^2. BMI values were converted to s.d. scores (BMI-SDS) that take into account the child's age and sex, using 1990 UK growth reference data[21] and computed with the programme lmsGrowth.[22] Reference values[21] were used to exclude implausible heights (<1.05 or >1.80 m), weights (<12 or >80 kg) and BMIs (<11 or >32).

Statistical analyses

All analyses were conducted on BMI-SDS that had been residualised for age and sex effects using a regression procedure. We used the GCTA package[23] to quantify the proportion of variance in BMI-SDS explained by 1.7 million SNPs for the unrelated children with genotype and BMI-SDS data. All possible pairs from a sample of 2269 individuals yields nearly 2.6 million pairwise comparisons (2 573 046). No pairs exceeded the GCTA standard cutoff coefficient of 0.025 for genetic relatedness, confirming that no two children in the analysed sample appeared to be genetically related in the traditional sense. We performed standard ACE model-fitting analyses using OpenMx[24] to estimate the heritability of BMI-SDS for the same sample of children by including anthropometric data from their co-twin to provide a direct comparison for the estimate derived from GCTA. The fit of the model was not of primary interest in this study; however, to assure a 'good fit', we used the full ACE Cholesky Decomposition Model (including additive genetic (A), shared environmental (C) and unique environmental (E) components), which is the full model and thus fits the data best and also provides estimates of the A, C and E parameters.

RESULTS

Of the 3152 children with genotyping data, 2402 families (76%) had provided anthropometric data and recorded age when the children were measured. Data from four children were excluded because they were reported as being <8 years old at the time they were measured; and 80 data points were excluded for implausible anthropometric results. Data from 22 children whose zygosity was unknown were excluded from the analyses because they could not be included in the twin analyses, along with a small number of children ($n=27$) with severe medical problems. Following exclusions, 2269 children had genotyping and anthropometric data.

The sample characteristics for the children included in the GCTA analysis are shown in Table 1. The average age was 9.9 years, 53% were girls and 39% were from monozygotic (identical) twin pairs. Their average BMI-SDS placed them close to the 1990 reference value, with comparatively low rates of overweight (8.7%) and obesity (3.6%).

The twin estimate of heritability of BMI-SDS in the sample was 82% (95% confidence interval: 0.74–0.88; $P<0.001$). Full results from the twin modelling are available from the first author. The GCTA estimate of genetic influence due to the additive effect of common SNPs was 30% (95% confidence interval 0.02–0.59; $P=0.02$). SNP heritability was therefore equivalent to 37% of the twin-estimated heritability (30%/82%). Figure 1 plots the variance explained in BMI-SDS from the twin analyses and the GCTA.

DISCUSSION

This is the first pediatric study to use GCTA to estimate the genetic influence on adiposity attributable to additive genetic effects from common SNPs across the entire genome. Consistent with findings in adults[11,13] the GCTA method gave an order-of-magnitude

Table 1. Characteristics of the GCTA analysis sample ($n=2269$ children)

	Mean (s.d.) or n *(%)*
Age (years)	9.90 (0.84)
Sex	
Females	1208 (53.2)
Males	1061 (46.8)
Zygosity	
Monozygotic	890 (39.2)
Dizygotic	1379 (60.8)
Weight (kg)	33.28 (7.33)
Height (m)	1.39 (0.08)
BMI (kg m^{-2})	17.03 (2.59)
BMI-SDS	−0.02 (1.13)
Weight status	
Healthy weight	1991 (87.7)
Overweight	197 (8.7)
Obese	81 (3.6)

Abbreviations: BMI, body mass index; BMI-SDS, BMI s.d. scores.

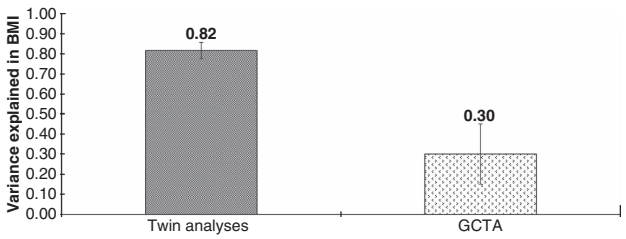


Figure 1. Comparison of variance explained in BMI (and s.e.) by genetic influences from twin analyses and GCTA at age 10.

increase compared with the GWAS estimates (1.5%), explaining 30% of the variance in BMI-SDS. This equated to 37% of the estimate of heritability derived from the twin design (82%) in the same families, and is comparable to many estimates derived from family studies.[3] Given that the 32 SNPs of the largest effect account for only 5% (1.5/30%) of the total common additive genetic SNP variance, these results suggest that 95% of the variation due to common SNPs have been undetected through GWAS. There are therefore likely to be hundreds of additional causal variants influencing childhood adiposity.

This GCTA estimate is likely to be at the lower end of the true additive genomic influence because it is limited to SNPs with a minor allele frequency of $\geqslant$5%; rarer variants are therefore excluded. In addition, causal variants that were not genotyped or not highly correlated with the SNPs on the genotyping array will also have been missed.

The GCTA value (30%) was larger than has been reported in studies in adults (10–16.5%),[11,13] suggesting that the additive genetic effect from common SNPs on BMI may be slightly higher for children. This is consistent with the higher estimates of heritability from pediatric than adult twin studies.[3,13] This may reflect the fact that adults are more likely than children to be making deliberate attempts at weight control, thus, limiting the observed genetic effect. The larger value may also be explained by the narrow age range of the sample, which reduces the effect of gene by age interaction.

These results have clinical and public health implications. Although the method used in the GCTA analysis cannot be used to predict obesity risk for any one individual because the genetic variants involved are not identified, the results underline the importance of additive genetic effects in the development of adiposity in childhood. This supports the current convention of using parental weight status as a proxy for childhood obesity risk.[25] Targeting children of obese parents for early-life obesity-prevention interventions, given that these children are most at risk, might be a useful direction to take.

This study has limitations. BMI tends to be lower in twins than singletons[26] and consistent with this, the average BMI of the sample placed them close to the 1990 reference value, and therefore below contemporary levels of adiposity. The sample size meant that it was not possible to carry out subgroup analyses. Height and weight data were parent-reported, therefore may be less reliable than researcher-measured anthropometrics, although they were found to be reliable in a subsample of families.[20] Lastly, the effects of pubertal status were not examined in this study; differences in pubertal status may have resulted in a slightly lower GCTA estimate of additive genetic effects on BMI-SDS. The study's strengths included the opportunity to estimate heritability using the twin design in the same sample for which we carried out the GCTA analysis.

These results find that GCTA analysis explains a substantial proportion of the genetic effect identified as 'missing heritability'. They provide compelling evidence that additive genetic influence from multiple common SNPs is a powerful determinant of adiposity in childhood.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The TEDS is supported by a programme grant to Robert Plomin from the UK Medical Research Council (G0901245, and previously G0500079), with additional support from the US National Institutes of Health (HD044454 and HD046167). The anthropometric data were collected as part of a grant to Jane Wardle from the UK Biological and Biotechnology Research Council (D19086), with additional support from programme grant funding from Cancer Research United Kingdom. Genome-wide genotyping was made possible by grants from the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z).

REFERENCES

1 Keiller SM, Colley JR, Carpenter RG. Obesity in schoolchildren and their parents. *Ann Hum Biol* 1979; **6**: 443–455.
2 Garn SM, Cole PE, Bailey SM. Living together as a factor in family-line resemblances. *Hum Biol* 1979; **51**: 565–587.
3 Elks CE, den HM, Zhao JH, Sharp SJ, Wareham NJ, Loos RJ *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne)* 2012; **3**: 29.
4 Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* 1997; **27**: 325–351.
5 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249 796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937–948.
6 Zhao J, Bradfield JP, Zhang H, Sleiman PM, Kim CE, Glessner JT *et al.* Role of BMI-associated loci identified in GWAS meta-analyses in the context of common childhood obesity in European Americans. *Obesity (Silver Spring)* 2011; **19**: 2436–2439.
7 Bradfield JP, Taal HR, Timpson NJ, Scherag A, Lecoeur C, Warrington NM *et al.* A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet* 2012; **44**: 526–531.
8 Elks CE, Loos RJ, Hardy R, Wills AK, Wong A, Wareham NJ *et al.* Adult obesity susceptibility variants are associated with greater childhood weight gain and a faster tempo of growth: the 1946 British Birth Cohort Study. *Am J Clin Nutr* 2012; **95**: 1150–1156.
9 Hebebrand J, Volckmar AL, Knoll N, Hinney A. Chipping away the 'missing heritability': GIANT steps forward in the molecular elucidation of obesity - but still lots to go. *Obes Facts* 2010; **3**: 294–303.
10 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**: 565–569.
11 Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011; **43**: 519–525.
12 Hunt SC, Hasstedt SJ, Kuida H, Stults BM, Hopkins PN, Williams RR. Genetic heritability and common environmental components of resting and stressed blood pressures, lipids, and body mass index in Utah pedigrees and twins. *Am J Epidemiol* 1989; **129**: 625–638.
13 Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet* 2012; **8**: e1002637.
14 Haworth CM, Davis OS, Plomin R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res Hum Genet* 2012; **16**: 117–125.
15 Trzaskowski M, Eley TC, Davis OS, Doherty SJ, Hanscombe KB, Meaburn EL *et al.* First genome-wide association study on anxiety-related behaviours in childhood. *PloS one* 2013, In Press.
16 Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, Phillips A *et al.* Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet* 2009; **41**: 1330–1334.
17 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
18 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
19 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; **2**: e190.
20 Wardle J, Carnell S, Haworth CM, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *Am J Clin Nutr* 2008; **87**: 398–404.

21 Cole TJ, Freeman JV, Preece MA. Body-mass index reference curves for the UK, 1990. *Arch Dis Child* 1995; **73**: 25–29.
22 Cole TJ. Software for LMS method: LMSGrowth PC 2009, Available at http://www.ucl.ac.uk/ich/research-ich/mrc-cech/research/studies/LMS-method2009; (accessed on 21 November 2012).
23 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
24 Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T *et al.* OpenMx: an open source extended structural equation modeling framework. *Psychometrika* 2011; **76**: 306–317.
25 Whitaker KL, Jarvis MJ, Beeken RJ, Boniface D, Wardle J. Comparing maternal and paternal intergenerational transmission of obesity risk in a large population-based sample. *Am J Clin Nutr* 2010; **91**: 1560–1567.
26 Pietilainen KH, Kaprio J, Rissanen A, Winter T, Rimpela A, Viken RJ *et al.* Distribution and heritability of BMI in Finnish adolescents aged 16y and 17y: a study of 4884 twins and 2509 singletons. *Int J Obes Relat Metab Disord* 1999; **23**: 107–115.

CHAPTER 5: FROM MODELLING TO MEASUREMENT: DEVELOPMENTAL TRENDS IN GENETIC INFLUENCE ON ADIPOSITY IN CHILDHOOD

This chapter has been adapted from a manuscript that is under review with Obesity:

Clare H Llewellyn*, Maciej Trzaskowski*, Robert Plomin and Jane Wardle

* These authors contributed equally

Running title: genes, development and obesity

ABSTRACT

Introduction: Twin-estimated heritability of BMI increases over childhood, despite high longitudinal genetic correlations; suggesting the same genes exert progressively stronger influence. Twin-derived statistics are inferential, however. In this study we use twin analyses, genome-wide complex trait analysis (GCTA), and an obesity-related polygenic risk score (PGS) in the same sample of children to better understand increasing genetic expression in BMI during childhood.

Methods: Data were from 2556 families from the Twins Early Development Study. BMI standard deviation scores (BMI-SDS) were calculated from validated parent-reported anthropometrics at 4 and 10 years using UK 1990 reference data. A longitudinal twin analysis estimated BMI-SDS heritability at 4 and 10, as well as the longitudinal genetic correlation. One child per family was selected at random for genotyping. Longitudinal GCTA used 1.7 million SNPs to determine DNA-based heritability for age 4 and 10, and the age-related genetic correlation. A polygenic obesity risk score (PGS) was created by summing 28 common obesity variants; bootstrapping was used to test the increase in the association between the PGS and BMI-SDS from 4 to 10 years.

Results: As expected, twin-estimated heritability was significantly higher at 10 (0.82; 95% confidence interval: 0.74-0.88) than 4 (0.43; 95% CI: 0.35-0.53). The PGS explained significantly more variance at 10 than 4 (R^2 Δ:0.024; 95% CI: 0.002-0.078). GCTA-estimated heritability did not increase significantly, but went from non-significant at 4 (0.20; 95% CI: -0.21-0.61) to significant at 10 (0.29; 95% CI: 0.01-0.57). Point estimates for the twin- (0.58) and GCTA-derived (0.66) longitudinal genetic correlations were of similar magnitude.

Conclusion: A high longitudinal genetic correlation from the GCTA supported the twin model in demonstrating that a large proportion of the measured additive common genetic variants that influence adiposity at age 4 also influence adiposity at age 10. The PGS-analyses suggested that the increasing heritability observed in the

twin analyses could be entirely explained by the same genes have increasing effects from early to late childhood.

INTRODUCTION

Body mass index (BMI) is a substantially heritable trait; with moderate-to-high estimates reported from family (24-81%) (1), adoption (20-60%) (2) and twin (47-90%) (1) studies. Using longitudinal data from the Twins Early Development Study (TEDS) we demonstrated that BMI increases significantly from 4 (49%) to 10 (82%) years of age (3). Similar developmental increases in heritability have been confirmed through meta-analysis for many other phenotypes (4). Increases in heritability are observed in combination with high genetic correlations over time; indicating that many of the genes that influence BMI at the later age are the same as the genes that influence BMI at the earlier age, but with greater genetic expression. However, twin-based statistics are inferential, relying on many assumptions.

Genome-wide association studies (GWAS) have now identified more than 32 common single nucleotide polymorphisms (SNPs) that are robustly associated with BMI in both adults and children (5, 6). These can be combined into a polygenic obesity risk score (PGS) to explore age-related change in the association between measured weight-related genes and BMI, in effect constraining the genetic correlation to be 1. Using a sub-set of these, studies have reported age-related increases in the association between obesity-related PGSs and BMI from early to late childhood using data from the Avon Longitudinal Study of Parents and Children (ALSPAC) (7) and from the 1946 British Birth Cohort Study (8). Directly comparing the association between an obesity-related PGS and BMI at different ages, with twin-based heritability in the same sample of children can shed light on the extent to which the increases in heritability can be explained by increasing genetic expression of the same genes.

A novel quantitative method called Genome-wide Complex Trait Analysis (GCTA) uses whole genome arrays to estimate the total additive genetic influence due to all common SNPs simultaneously. Like twin data, GCTA can

be used to explore developmental increases in genetic expression insofar as it provides age-specific estimates of the additive effects of all common SNPs, as well as an estimate of the longitudinal genetic correlation between those genetic influences at the two time points. GCTA therefore provides quantitative estimates of genetic effects using measured genes, to complement the inferential twin-based estimates. Longitudinal GCTA has never been used to explore developmental change in genetic expression of BMI Combining longitudinal GCTA- and twin-based models of genetic effects on BMI in the same sample of children makes it possible to substantiate inferential twin statistics with DNA-based estimates.

In this study we take a novel approach by combining GCTA, associations with a PGS, and twin analyses to explore age-related increases in BMI in the same sample of children in order to test the hypothesis that developmental increases in heritability are driven primarily by the same genes exerting stronger effects as children get older.

METHODS

SAMPLE

The sample for this study was recruited from the Twins Early Development Study (TEDS), a large population-based cohort of >11,000 British twins (9). Informed consent for each part of the study was provided by both the children and their parents prior to data collection. Ethical approval was provided by King's College London's Ethics Committee.

GENOTYPING

In 2010, genome-wide genotyping was carried out for one randomly selected child in each of 3665 families as part of the WTCC2 project (10). DNA was extracted from buccal cheek swabs; the Affymetrix 6.0 GeneChip was

used to genotype about 1,000,000 SNPs using standard experimental protocols. The IMPUTE version 2 software (11) was used to impute approximately 2.5 million additional SNPs. After stringent quality control carried out as part of WTCCC2 project, the sample was reduced to about 1.7 million SNPs and 3152 children.

CREATING A POLYGENIC RISK SCORE

A polygenic obesity risk score (PGS) indexing genetic predisposition to obesity was calculated using 28 known obesity SNPs from two published meta-analyses in adults (6) and children (5). 24 of the 34 obesity-risk increasing SNPs were available on the Affymetrix 6.0 GeneChip: rs9939609 (*FTO* (6)), rs2867125 (*TMEM18* (6)), rs571312 (*MC4R* (6)), rs10938397 (*GNPDA2* (6)), rs10767664 (*BDNF* (6)), rs2815752 (*NEGR1* (6)), rs7359397 (*SH2B1* (6)) rs3817334 (*MTCH2* (6)), rs29941 (*KCTD15* (6)), rs543874 (*SEC16B* (6)), rs987237 (*TFAP2B* (6)), rs7138803 (*FAIM2* (6)), rs10150332 (*NRXN3* (6)), rs713586 (*RBJ; POMC* (6)), rs12444979 (*GPRC5B* (6)), rs2241423 (*MAP2K5* (6)), rs1514175 (*TNNI3K* (6)), rs10968576 (*LRRN6C* (6)), rs887912 (*FANCL* (6)), rs13078807 (*CADM2* (6)), rs1555543 (*PTBP2* (6)), rs206936 (*NUDT3* (6)), rs9568856 (*OLFM4* (5)), rs9299 (*HOXB5* (5)). It was possible to index 4 of the 34 obesity-risk increasing SNPs using proxy SNPs that were in high linkage disequilibrium with the original SNP (R^2>0.9): rs2112347 (*FLJ35779*) was indexed using rs3797580 (R^2=1); rs4836133 (*ZNF608*) was indexed using rs6864049 (R^2=1); rs4929949 (*RPL27A*) was indexed using rs9300093 (R^2=0.97); rs3810291 (*TMEM160*) was indexed using rs7250850 (R^2=1). Six of the 34 obesity-risk increasing SNPs could not be measured directly or reliably tagged using proxy SNPs: rs2890652 (LRP1B), rs9816226 (ETV5), rs13107325 (SLC39A8), rs4771122 (MTIF3), rs11847697 (PRKD1), rs2287019 (QPCTL).

A PGS was created by calculating a mean score for each participant from the 24 available obesity-risk increasing SNPs, and the 4 proxy SNPs. The possible score ranged from 0-56 with higher scores indicating a greater genetic predisposition to obesity. A weighted mean score was then calculated to take account of the relative effect size of each SNP; each SNP was multiplied by its beta coefficient published in the meta-analyses from which the obesity-risk increasing SNPs were identified (5, 6).

MEASUREMENT OF BMI-SDS AT AGE 4 AND 10 YEARS

Anthropometric data were collected from the children in 1999 and 2005, when children were 4 years and 10 years old, respectively. Questionnaires and 2-meter tape measures were mailed to the families; parents were provided with detailed instructions regarding how to measure their children's height (to the nearest centimeter) and weight (to the nearest pound or tenth of a kilogram), and they were asked to record the date of each measurement. Correspondence between parent- and researcher-measured height and weight were 0.90 and 0.83 in a subsample of 228 families (12).

Body mass index (BMI) was calculated from height and weight (weight (kg)/ height $(m)^2$) and converted to Standard Deviation Scores (BMI-SDS) which take into account the child's age and sex, using 1990 UK growth reference data (13). BMI-SDS were computed using the program ImsGrowth (14). Reference data (13) were used to exclude implausible anthropometric values (age 4: heights <0.80m or >1.35m; weights <9 kg or >40 kg; BMIs <12 or >25; age 10: heights <1.05m or >1.80m; weights <12 kg or >80 kg; BMIs <11 or >32), as well as outliers for the age at which anthropometrics were measured (age 4 data: $\leq$ 2 years or $\geq$ 5.99 years; age 10 data: $\leq$ 8 years or $\geq$ 11.99 years). BMI-SDS were residualized for age-and sex-effects of the sample using a regression procedure, prior to analyses.

EXCLUSIONS

2556 children of the sample with genotyping data (n=3152) had provided height and weight measurements for at least one of the two time periods, and reported the exact age of the child when they were weighed and measured. As the sample was selected from GWA discovery individuals and the GWA sample was already screened for medical and other general exclusions, there were no further exclusions (for details on that sample please see (15)).

Associations between Polygenic Risk Score and BMI-SDS at ages 4 and 10 years

Linear regression analyses were used to establish the association between the PGS with age- and sex-adjusted BMI-SDS at 4 and 10 years of age for the 2556 unrelated children with genotyping and anthropometric data. Bootstrapping analyses were carried out to provide 95% confidence intervals for the R^2 estimates derived from the linear regression analyses, and to test for differences between the R^2 estimated at age 4 and age 10 (Supplementary Figure 1). To provide a more accurate test for bootstrapping the difference in R^2 between the two ages, we sampled from individuals who had data-points at both ages. The linear regression and bootstrapping analyses were performed in R version 2.15.0 (16).

Genome-wide Complex Trait Analysis (GCTA) at ages 4 and 10 years

GCTA takes advantage of existing GWA data to estimate the total amount of a trait's variance that can be explained by the additive effects of all common SNPs measured on commercial chips, and SNPs in high linkage disequilibrium (LD) with them. The method adapts a linear mixed model (LMM) framework fitting genetic influence as a random polygenic effect. Although GCTA is the name of a package developed to perform this adapted LMM, the method does not have an established name so we use the name of the package to refer to the method itself in order to improve the legibility. Conceptually, GCTA estimates the amount of genetic influence by associating mean genetic similarity calculated from all genetic loci to the phenotypic similarity between all pairs of unrelated individuals in the sample (17). Using a random effects model to estimate genetic influence means that this estimate is 'unbiased' when all individuals are truly unrelated. For that reason, if any pairwise comparison returns relatedness higher than a fourth cousin (genetic relatedness > 0.025), one of the pair is removed. Given that GCTA is a genome-wide method, it is affected by population structure (18), like GWA. To account for this we used eight eigen vectors previously used on the same sample in our GWAS (15).

To compare age-related differences in the amount of variance in BMI-SDS explained by the additive effects of common SNPs we used a bivariate GCTA model (19); this differs to a univariate model insofar as it focuses on the covariance between two time-points rather than on the variance of a single time-point. It provides age-specific estimates of heritability due to additive effects of common SNPs, as well as a genetic correlation indicating the extent to which the additive effects of common SNPs on BMI-SDS at 4 and 10 are the same.

Twin analyses of the heritability of BMI-SDS at ages 4 and 10 years

In order to provide twin-based heritability of BMI at ages 4 and 10 for comparison, data from each child's co-twin was modeled using a longitudinal Cholesky Decomposition Model in OpenMx (20). The fit of the model was not of primary interest in this study; however, to assure a 'good fit', we used the full model with all parameters, including additive genetic (A), shared environmental (C), and unique environmental (E) components of variance for each age, and allowed longitudinal covariation among all of these parameters. Because previous analysis on the same sample did not indicate sex differences we modeled both sexes together.

RESULTS

SAMPLE CHARACTERISTICS

The sample of children included slightly more females (n=1368/2556; 54%) than males (n=1188; 46%), and the majority were from dizygotic (n=1535; 60%) rather than monozygotic (n=1021; 40%) twin pairs. The number of obesity risk-increasing alleles was normally distributed within the sample and ranged from 11 to 30 (mean=12.48; SD=2.85) (**Figure 1**). The weighted polygenic risk score ranged from -0.08 to 0.03 (mean=-0.03; SD=0.02). The anthropometric characteristics of the sample are shown in **Table 1** for ages 4 and 10. At each age the mean BMI-SDS was less than 0, indicating that on average the relative body weight of the sample was

slightly less than the UK reference values for those ages. In keeping with this, the sample had slightly lower rates of overweight and obesity at the two ages than typically observed in national UK statistics.

TWIN ANALYSES OF THE HERITABILITY OF BMI-SDS AT AGES 4 AND 10 YEARS

We confirmed the developmental increases in the heritability of BMI-SDS reported previously. Twin-estimated heritability increased significantly from 4 to 10 years, increasing from a moderate 0.43 (95% confidence interval: 0.35-0.53) at age 4 to 0.82 (95% confidence interval: 0.74-0.88) at age 10 (**Table 2**). The genetic correlation between BMI-SDS at ages 4 and 10 was 0.58 (95% confidence interval: 0.48-0.68). The full results from the twin analysis are available in online materials (Supplementary Table 2).

GCTA ESTIMATES OF THE HERITABILITY OF BMI-SDS AT AGES 4 AND 10 YEARS

The GCTA results were in keeping with the twin analyses, insofar as the point-estimate for the GCTA-derived genetic correlation was high, and in keeping with that derived from the twin model, but due to a large standard error it was not significant (r=0.66; 95% confidence interval: -0.28-1.60). The estimate of heritability increased from a non-significant 0.20 (95% confidence interval: -0.21-0.61) at age 4 to a significant 0.29 (95% confidence interval: 0.01-0.57) at age 10 (Table 2). Full GCTA results are available in online materials (Supplementary Table 1).

ASSOCIATIONS BETWEEN THE OBESITY-RELATED PGS AND BMI-SDS AT AGES 4 AND 10 YEARS

The analyses with the PGS also supported the twin-based statistics. In the multiple linear regression analyses the PGS was significantly associated with BMI-SDS at both ages, but the size of the association increased from age 4 (R^2=0.008; 95% confidence interval: 4.3e-09–0.042; p=0.002) to age 10 (R^2=0.034; 95% confidence interval: 0.009–0.093; p<0.001) (Table 2; Figure 1). Bootstrapping analyses confirmed that the association was

significantly higher at age 10 than at age 4 ($R^2\Delta$=0.025; bootstrapped 95% confidence interval =0.004-0.056). Directly comparing the PGS analyses with the twin-based heritability indicated that at age 4 the PGS accounted for 1.9% of the twin-estimated heritability (0.008/0.43), and at age 10 it accounted for 4% of the heritability, suggesting that increasing effects of the same genes could entirely explain the increases in twin-estimated heritability from age 4 to age 10.

DISCUSSION

In this study, we combined GCTA and an obesity-related PGS with twin analyses in the same sample of individuals to test the hypothesis that the developmental increases in the heritability of BMI-SDS are driven primarily by the same genes exerting increasing effects as children get older. In line with previous estimates we confirmed that twin-estimated heritability of BMI-SDS increased significantly from age 4 to age 10 ($h^2\Delta$=35%) (3). The point-estimates of twin- and GCTA-estimated genetic correlations were of equal magnitude (0.58 and 0.66 respectively), providing a DNA-based estimate in support of the inferential twin-based statistic and indicating that the increasing heritability is driven largely by the same weight-related genes exerting progressively greater effects. There was also a significant increase in the association between the obesity-related PGS and BMI-SDS from age 4 to age 10 ($R^2\Delta$=0.026). A direct comparison of PGS and twin analyses showed that the PGS explained 1.9% of twin-based heritability at age 4, and 4.1% of twin-based heritability at age 10; demonstrating that the increase in twin-estimated heritability could be explained entirely by the same genes exerting greater effects as children get older.

These findings support studies of both specific obesity-related genetic variants, and studies of PGSs. A meta-analysis of eight cohorts of European ancestry showed that the primary obesity-related common genetic variant in the FTO gene was only significantly associated with higher BMI after 5.5 years, after which time each obesity-risk allele increased BMI percentage change progressively from 5.5-7 years (0.7%) to 7-9 years (1.0%) to 9-11 years (1.3%) (21). Analyses of an obesity-associated variant in the MC4R gene showed a similar pattern

in the 1946 British Birth Cohort Study; associations with age- and sex-adjusted weight increased during childhood and adolescence by 0.005 units per year.

Studies of obesity-related PGSs have shown similar increases in associations with BMI from early to late childhood. In ALSPAC a PGS comprising 8 obesity-related SNPs showed only a weak association with BMI-SDS during infancy until 3.5 years, but from 3.5-11 years the size of the association increased rapidly (7). Lifecourse analyses have been conducted for a 29-SNP PGS in the Dunedin Multidisciplinary Health and Development Study (22), and for an 11-SNP PGS in the 1946 Birth Cohort (8); both studies showed that the association with BMI increased year on year from early to late childhood, in keeping with ALSPAC findings and the present analysis. ALSPAC and the Dunedin cohort reported age-specific effect sizes allowing for comparison with our study; the age-matched associations in TEDS were almost identical to those reported for the Dunedin cohort that used a comparable PGS, and were higher than estimates from ALSPAC that utilized fewer SNPs (TEDS at 4 years: 0.8%; Dunedin at 3 years: 0.6%; ALSPAC at 3.5 years: 0.2%; TEDS at 10 years: 3.4%; Dunedin at 9 years: 3.2%; ALSPAC at 10 years: 1.6%). Together, these studies provide convincing evidence that the same weight-related genes have progressively stronger effects from early to late childhood.

Only one previous study has conducted longitudinal GCTA and compared DNA-based longitudinal genetic correlations with twin-based estimates. This study explored developmental genetic expression for a range of cognitive abilities, also using TEDS. Increasing heritability with age was demonstrated by both GCTA and twins, in combination with high genetic correlations of comparable magnitude for each method (23). We had a small sample size for our BMI analyses at age 4 (n=1419), rendering the current study underpowered to detect a significant longitudinal genetic correlation. The findings need to be replicated using a larger sample.

Age-related increases in heritability have also been shown for externalizing behaviors, anxiety symptoms, depressive symptoms, IQ and social attitudes (4) from early adolescence into early adulthood (4). One explanation that has been put forward to account for this common observation is that active gene-environment correlations increase as children get older and gain independence (4). That is, early childhood

environments are largely controlled by parents, but as children grow older and gain independence they have more control over their behavior, and may increasingly select out environments that 'indulge' their genetic propensities. It has been hypothesized that one of the mechanisms through which genes exert their effects on weight is appetite regulation (24); and in particular, FTO's effects on childhood BMI have been shown to be mediated via satiety sensitivity (25, 26). A corollary of this is that FTO is unlikely to be expressed fully unless an individual has the freedom to consume to satiety – a privilege that comes with the increasing independence gained from early to middle childhood, as children have more freedom to make decisions with regard to how much and when to eat. Many of the other obesity-related SNPs are also hypothesized to influence weight via central appetite control processes (6), making gene-environment correlation a plausible explanation for this observation.

It is also possible that the increasing genetic effect on BMI from early to late childhood reflects processes that began even earlier in life. The Dunedin Multidisciplinary Health and Development Study showed that early life growth from birth to 3 years significantly mediated some of the genetic risk of obesity later on in adolescence (22). The study also reported an earlier adiposity rebound, and a higher BMI at adiposity rebound (22), for individuals at highest genetic risk of obesity.

These findings have public health implications; genetic predisposition to obesity is increasingly expressed from the time a child begins school at age 4, to the transition to secondary school. The beginning of school may therefore provide an important intervention window for public health initiatives aimed at reducing or preventing the development of childhood obesity. In support of this, a recent paper used data from the National Childhood Obesity Center in Sweden to show that the efficacy of long-term behavioral treatment for severely obese children is considerably more effective in young children (6-9 years) compared to older children (10-13 years) or adolescents (14-16 years) (27). The findings from the current study also have methodological implications insofar as they show that the genetic architecture underlying developmental increases captured by twin-data can be replicated using DNA alone.

This study has a number of strengths. It is the first study to combine twin analyses, GCTA and a PGS in the same sample of children to better understand age-related increases in BMI heritability. The study also had some limitations. The sample of children used in this analysis only included those with genotyping data as well as data for BMI at either 4 years or 10 years, or both years (n=2556). This was a small proportion of the overall sample. It is possible that parents of overweight and obese children were less willing to send in weight data, limiting the generalizability of the results. In relation to this, the sample mean approximated the 1990 reference value, and rates of overweight and obesity were relatively low, indicating that the analysis sample was somewhat leaner than current UK children of the same age.

In conclusion, in this study we combined twin-, GCTA- and PGS- analyses for the first time to show that the increases in the heritability of BMI-SDS from early to late childhood can be entirely explained by increasing effects of the same genes, substantiating previous inferential twin statistics. These findings underline the importance of intervening early in life for the prevention of childhood obesity.

ACKNOWLEDGEMENTS

The Twins Early Development Study (TEDS) is supported by a program grant to Robert Plomin from the UK Medical Research Council (G0901245; and previously G0500079), with additional support from the US National Institutes of Health (HD044454; HD046167). The anthropometric data were collected as part of a grant to Jane Wardle from the UK Biological and Biotechnology Research Council (D19086) with additional support from program grant funding from Cancer Research UK. Genome-wide genotyping was made possible by grants from the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z; 085475/Z/08/Z).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Elks CE, den HM, Zhao JH et al. Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne)*. 2012;3:29.

2. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet*. 1997;27(4):325-51.

3. Haworth CM, Carnell S, Meaburn EL, Davis OS, Plomin R, Wardle J. Increasing heritability of BMI and stronger associations with the FTO gene over childhood. *Obesity (Silver Spring)*. 2008;16(12):2663-8.

4. Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. *Twin Res Hum Genet*. 2007;10(3):423-33.

5. Bradfield JP, Taal HR, Timpson NJ et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet*. 2012;44(5):526-31.

6. Speliotes EK, Willer CJ, Berndt SI et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010;42(11):937-48.

7. Elks CE, Loos RJ, Sharp SJ et al. Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med*. 2010;7(5):e1000284.

8. Elks CE, Loos RJ, Hardy R et al. Adult obesity susceptibility variants are associated with greater childhood weight gain and a faster tempo of growth: the 1946 British Birth Cohort Study. *Am J Clin Nutr*. 2012;95(5):1150-6.

9. Haworth CM, Davis OS, Plomin R. Twins Early Development Study (TEDS): A Genetically Sensitive Investigation of Cognitive and Behavioral Development From Childhood to Young Adulthood. *Twin Res Hum Genet*. 2012;1-9.

10. Barrett JC, Lee JC, Lees CW et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet*. 2009;41(12):1330-4.

11. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.

12. Wardle J, Carnell S, Haworth CM, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *Am J Clin Nutr*. 2008;87(2):398-404.

13. Cole TJ, Freeman JV, Preece MA. Body-Mass Index Reference Curves for the Uk, 1990. *Archives of Disease in Childhood*. 1995;73(1):25-9.

14. Cole, T. J. Software for LMS method: LMSGrowth PC. 2009.

15. Trzaskowski M, Eley TC, Davis OS et al. First genome-wide association study on anxiety-related behaviours in childhood. *PloS one*. 2013;8(4):e58676.

16. R Development Core Team. R: A language and environment for statistical computing. 2012. Vienna, Austria, R Foundation for Statistical Computing.

17. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565-9.

18. Yang J, Weedon MN, Purcell S et al. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet*. 2011;19(7):807-12.

19. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. 2012;28(19):2540-2.

20. Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T. OpenMx: an open source extended structural equation modeling framework. *Psychometrika*. 2011;76:306-17.

21. Sovio U, Mook-Kanamori DO, Warrington NM et al. Association between Common Variation at the FTO Locus and Changes in Body Mass Index from Infancy to Late Childhood: The Complex Nature of Genetic Association through Growth and Development. *PLoS Genet*. 2011;7(2):e1001307.

22. Belsky DW, Moffitt TE, Houts R et al. Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study. *Arch Pediatr Adolesc Med*. 2012;166(6):515-21.

23. Trzaskowski M, Davis OS, DeFries JC, Yang J, Visscher PM, Plomin R. DNA Evidence for Strong Genome-Wide Pleiotropy of Cognitive and Learning Abilities. *Behav Genet*. 2013.

24. Llewellyn CH, van Jaarsveld CH, Plomin R, Fisher A, Wardle J. Inherited behavioral susceptibility to adiposity in infancy: a multivariate genetic analysis of appetite and weight in the Gemini birth cohort. *Am J Clin Nutr*. 2012;95(3):633-9.

25. Wardle J, Carnell S, Haworth CM, Farooqi IS, O'Rahilly S, Plomin R. Obesity associated genetic variation in FTO is associated with diminished satiety. *J Clin Endocrinol Metab*. 2008;93(9):3640-3.

26. Wardle J, Llewellyn C, Sanderson S, Plomin R. The FTO gene and measured food intake in children. *Int J Obes (Lond)*. 2009;33(1):42-5.

27. Danielsson P, Kowalski J, Ekblom O, Marcus C. Response of severely obese children and adolescents to behavioral treatment. *Arch Pediatr Adolesc Med*. 2012;166(12):1103-8.

Table 1 Summary statistics and anthropometrics for the analysis sample at age 4 and 10 years (n=2556 children)[a]

	Mean (sd) or n (%)	
	Age 4	**Age 10**
Age at measurement (years)	4.00 (0.11)	9.90 (0.85)
Weight (kg)	16.53 (2.40)	33.46 (7.78)
Height (m)	1.02 (0.05)	1.39 (0.08)
BMI[b] (kg/m^2)	15.79 (2.10)	17.03 (2.57)
BMI-SDS[c]	-0.12 (1.59)	-0.02 (1.12)
Weight status[d]		
Healthy weight	1270 (87.1)	2021 (87.6)
Overweight	94 (6.4)	201 (8.7)
Obese	94 (6.4)	86 (3.7)

[a] The sample characteristics presented are for the 2556 unrelated children with genotyping and BMI-SDS data for at least one age point; 1419 unrelated children had genotyping and BMI-SDS data at age 4; 2268 children had genotyping and BMI-SDS data at age 10.

[b] BMI, body mass index.

[c] BMI-SDS, BMI standard deviation score: BMI adjusted for age and sex using UK 1990 reference data (13).

[d] Weight status calculated from BMI-SDS using UK 1990 reference data: healthy weight, BMI-SDS <91st centile; overweight, BMI-SDS ≥91st centile and <98th centile; obese, BMI-SDS ≥98th centile (13).

Table 2. Comparison of twin-, GCTA- and PGS-estimates of heritability at 4 and 10 years

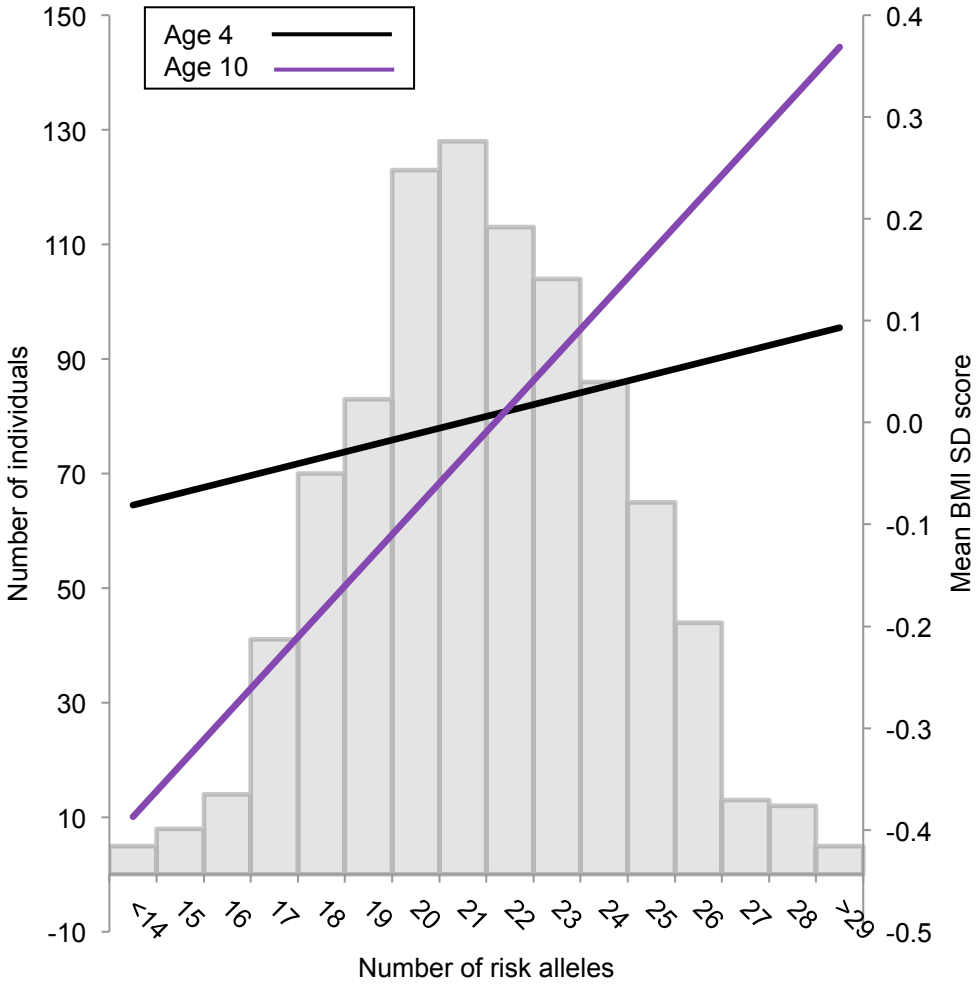
	Twin study estimate of heritability (95% confidence interval)	**GCTA estimate of genetic influence (95% confidence interval)**	**Association (R2) between PGS and BMI-SDS (95% confidence interval)[b]**
BMI-SDS age 4	**0.43** (0.35-0.53) [a]	**0.20 (-0.21 – 0.61)**	**0.008** (4.3e-09 - 0.042) [c]
BMI-SDS age 10	**0.82** (0.74-0.88) [a]	**0.29 (0.01 – 0.57)**	**0.034** (0.009-0.093) [c]

[a] Twin-estimated heritability was significantly higher at age 10 compared to age 4.

[b] Bootstrapping was used to test the difference in the association (R2Δ) between the PGS and age- and sex-adjusted BMI-SDS at 4 and 10 years.

[c] R^2 estimates at 4 and 10 years were significantly different from one another (R2Δ=0.024, 95% CI=0.002-0.078).

Figure 1. Regression of mean age- and sex-adjusted BMI-SDS at 4 and 10 years (n=2556)



CHAPTER 6: DNA EVIDENCE FOR STRONG GENETIC STABILITY AND INCREASING HERITABILITY OF INTELLIGENCE FROM AGE 7 TO 12

This chapter is presented as a published paper and is an exact copy of the following journal publication:

Maciej Trzaskowski, Jin Yang, Peter M. Visscher, Robert Plomin (2013). *Molecular Psychiatry*.

OPEN

Molecular Psychiatry (2013), 1–5

www.nature.com/mp

ORIGINAL ARTICLE

DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12

M Trzaskowski[1], J Yang[2], PM Visscher[2,3] and R Plomin[1]

Two genetic findings from twin research have far-reaching implications for understanding individual differences in the development of brain function as indexed by general cognitive ability (g, aka intelligence): (1) The same genes affect g throughout development, even though (2) heritability increases. It is now possible to test these hypotheses using DNA alone. From 1.7 million DNA markers and g scores at ages 7 and 12 on 2875 children, the DNA genetic correlation from age 7 to 12 was 0.73, highly similar to the genetic correlation of 0.75 estimated from 6702 pairs of twins from the same sample. DNA-estimated heritabilities increased from 0.26 at age 7 to 0.45 at age 12; twin-estimated heritabilities also increased from 0.35 to 0.48. These DNA results confirm the results of twin studies indicating strong genetic stability but increasing heritability for g, despite mean changes in brain structure and function from childhood to adolescence.

Molecular Psychiatry advance online publication, 29 January 2013; doi:10.1038/mp.2012.191

Keywords: cognitive development; Genome-wide complex trait analysis (GCTA); intelligence; twins

INTRODUCTION

Although developmental research from childhood to adolescence reveals species-general changes in brain structure and function,[1,2] much less is known about the development of individual differences within our species, which has been called 'one of the preeminent challenges of neuroimaging'.[3] It is important to understand the developmental etiology of individual differences, because societal problems often involve individual differences—for example, why some children are slow to speak, to learn or to read. The description and causes of species' means are not necessarily related to the description and causes of variances within a species.[4] Two well-replicated genetic findings from twin studies comparing monozygotic and dizygotic (DZ) twins suggest hypotheses at the level of individual differences in cognitive ability that may be relevant to neuroscience, to the extent that brain structure and function underlie cognitive outcomes. These twin-study findings involve general cognitive ability, which was labeled *g* by Spearman more than a century ago,[5] but is commonly known as *intelligence*.[6] g is the most researched cognitive trait in genetics[7] and has important links with neuroscience.[8,9]

First, the heritability of g increases during development, even from childhood to adolescence.[10] This finding is counterintuitive to the extent that genetic effects are thought to be static, and environmental effects are expected to accumulate during development. The increasing heritability of g also seems at odds with the second genetic finding: The same genes largely affect g throughout development.[11] For example, in a longitudinal twin analysis from childhood to adolescence, the genetic correlation was estimated as 0.96, although the 95% confidence interval for this estimate was 0.74–1.0.[12] The genetic correlation is literally the correlation between genetic effects on g at the two ages independent of heritability.[11] The high genetic correlation implies that if a gene is found to be associated with g in childhood, the gene is also highly likely to be associated with g in adolescence. Later, we offer a hypothesis as to how heritability can increase when genetic effects are stable from age to age.

These two genetic findings have not found much traction in the neurodevelopmental literature. This neglect might be due in part to a lack of attention to individual differences, but it might also be due to skepticism about the twin method, which relies on some major assumptions, most notably, equal environmental treatment of monozygotic and DZ twins.[11] Quantitative genetic designs such as the twin method would no longer be needed if it were possible to identify all of the genes responsible for heritability.[13] However, it has proven more difficult than expected to identify genes for complex traits,[14] including g,[15] which has led to the refrain of 'missing heritability'.[16,17] Nonetheless, it is now possible to use DNA itself to estimate genetic variance and covariance in any sample of unrelated individuals, not just samples consisting of special family members such as twins or adoptees. The method, called genome-wide complex trait analysis (GCTA)[18] correlates genomic similarity across hundreds of thousands of single nucleotide polymorphisms (SNPs) with phenotypic similarity in a large sample of unrelated individuals.[19] This population-based DNA approach does not rely on the strong assumptions made in classical twin studies. GCTA compares similarity across hundreds of thousands of SNPs with phenotypic similarity pair by pair in a large sample of unrelated individuals. Although conventionally unrelated individuals only vary in their genetic similarity by a small amount, GCTA accumulates all the genotype – phenotype association signals using the massive information available in a matrix of thousands of individuals, each compared pair by pair with every other individual in the sample. GCTA has been used to

[1]King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Denmark Hill, London, UK; [2]University of Queensland Diamantina Institute, The University of Queensland, Princess Alexandra Hospital, Brisbane, Queensland, Australia and [3]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia. Correspondence: M Trzaskowski, MSc, King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, PO80, Institute of Psychiatry, DeCrespigny Park, Denmark Hill, London SE5 8AF, UK.
E-mail: maciej.trzaskowski@kcl.ac.uk
Received 16 October 2012; accepted 12 November 2012

estimate genetic influence for height,[19] weight,[20] psychiatric and medical disorders,[21–23] personality[24] and even economic and political preferences.[25] GCTA has also been applied to g in adults[26] and children.[27] These GCTA estimates of genetic influence, although substantial, have been lower than heritability typically found in twin studies of these traits. Using the 12-year data from the sample in the present report, GCTA and twin estimates of heritability were compared explicitly for several cognitive measures; the GCTA estimate of g was 35% and the twin estimate was 46%.[28] Precision in comparing GCTA and twin estimates is important because, as explained later, this comparison reveals important information about a trait's genetic architecture.

This previous GCTA research involves univariate analysis in that it decomposes the phenotypic variance of a single trait into genetic and non-genetic components of variance. Recently, GCTA has been extended to bivariate analysis, which decomposes the phenotypic covariance between traits into components of covariance. The first preliminary attempt to extend GCTA to bivariate analysis reported a genetic correlation of 0.62 for g in childhood (age 11) and old age.[27] Here, we use a new bivariate GCTA method[18,29] to test the hypotheses of strong stability and increasing heritability for g from age 7 to 12. We also compare GCTA estimates with those from a twin analysis based on the same sample at the same ages using the same measures.

MATERIALS AND METHODS

Sample

The sample was drawn from the Twins Early Development Study (TEDS), which is a multivariate longitudinal twin-study that recruited over 11 000 twin pairs born in England and Wales in 1994, 1995 and 1996.[30,31] TEDS is representative of the UK population.[32] The project received approval from the Institute of Psychiatry ethics committee (05/Q0706/228), and parental consent was obtained before data collection. Individuals were included if their first language was English and they had no major medical or psychiatric problems. GCTA was conducted on g at ages 7 and 12 for 2875 unrelated individuals in TEDS (only one member of a twin pair), of which 1334 had g data at both ages. Twin model-fitting analyses of g at ages 7 and 12 were conducted for 6702 TEDS twin pairs, of which 2269 pairs had g data at both ages. As expected for representative twin studies, the twins included similar numbers of monozygotic twins, same-sex DZ twins and opposite-sex DZ twins.

Genotyping

Although DNA is available for more than 12 000 TEDS participants, funds were available to genotype 3665 individuals (one member only per twin pair) on Affymetrix GeneChip 6.0 (Affymetrix Inc., Santa Clara, CA, USA) SNP genotyping arrays using standard experimental protocols as part of the WTCCC2 project. In addition to nearly 700 000 genotyped SNPs, more than one million other SNPs were imputed using IMPUTE v.2 software (https://mathgen.stats.ox.ac.uk/impute/impute.html).[33] DNA for 3152 individuals (1446 males and 1706 females) survived quality control criteria. Of these 3152 individuals, 2875 had g scores at least at one age and 1344 had g scores at both ages. To control for ancestral stratification, we performed principal component analyses on a subset of 100 000 quality-controlled SNPs after removing SNPs in linkage disequilibrium ($r^2 > 0.2$).[34] Using the Tracy – Widom test,[35] we identified 8 axes with $P < 0.05$, which were used as covariates in GCTA analyses.

Measures

The measures and testing procedures have been described in detail for age 7[36] and 12.[37] At each age, a composite index of g was derived from two verbal tests and two non-verbal tests. At age 7, the two verbal tests consisted of the Similarities subtest and the Vocabulary subtest from the WISC-III-UK, and the two non-verbal tests were the picture completion subtest from the WISC-III-UK and the Conceptual Grouping subtest from the McCarthy Scales of Children's Abilities. At age 12, the verbal tests included the Information and Vocabulary subtests from the WISC-III-PI Multiple Choice test, and the two non-verbal reasoning tests were WISC-III-UK Picture Completion and Raven's Standard and Advanced Progressive Matrices. At age 7, testing was conducted by telephone as described elsewhere;[36] at age 12, testing was conducted online.[37] For each cognitive measure at each age, scores were regressed on sex and age and standardized residuals were derived, ranked with random values given to tied data, and quantile normalized.[38,39] Finally, total composites for g were created as unit-weighted means requiring complete data for at least three of the four tests. All the procedures were executed using R (www.r-project.org).[40]

Statistical analyses

Genome-wide complex trait analysis. The first step in GCTA is to calculate pairwise genomic similarity between all pairs of individuals in the sample using all genetic markers genotyped on the SNP array. Because GCTA is designed to estimate genetic variance due to linkage disequilibrium between unknown causal variants and genotyped SNPs from a sample of unrelated individuals in the population, any close genetic relatedness is eliminated; for this reason any individual whose genetic similarity is equal to or greater than a fourth cousin is removed (estimate of pairwise relatedness > 0.025). The essence of GCTA is to compare a matrix of pairwise genomic similarity to a matrix of pairwise phenotypic similarity using a random-effects mixed linear model.[18] In univariate analysis, the variance of a trait can be partitioned using residual maximum likelihood into genetic and residual components. Detailed description of this method can be found in Yang, Lee *et al.*[18] and Yang, Benyamin *et al.*[19] The bivariate method extends the univariate model by relating the pairwise genetic similarity matrix to a phenotypic covariance matrix between traits 1 and 2, allowing for correlated residuals.[29] The eight principal components described earlier were used as covariates in our GCTA analyses; as mentioned, all phenotypes were age- and sex-regressed before analysis.

Twin modeling. The classical twin design and model-fitting is discussed elsewhere.[11] We fit a bivariate twin model using OpenMx,[41] which provided a direct comparison with the bivariate GCTA. The correlated factor solution is the least restricted model allowing variables to correlate with one another via genetic, shared environment and non-shared environment. Because previous analyses of these data indicated nonsignificant differences in model-fitting results between males and females,[32,42] we combined same-sex and opposite DZ twin pairs in order to increase the power of the analyses. Twin analyses limited to same-sex twins yielded highly similar results (available from the first author).

RESULTS AND DISCUSSION

Genetic stability

As shown in Table 1, the GCTA genetic correlation between g at ages 7 and 12 was 0.73 (0.29 standard error, s.e.). Table 2 shows that the twin-study yielded a highly similar genetic correlation of 0.75 (0.08 s.e.). The genetic correlation indexes the correlation between genetic effects on g at the two ages independent of heritability. That is, the genetic correlation can be high even if heritability is low. It is also possible to weight the genetic correlation by heritability in order to estimate the genetic contribution to the phenotypic correlation. The phenotypic correlation for g between ages 7 and 12 was 0.46 (0.02) for 2408 children (one member randomly chosen from each twin pair) with g data at both ages. For GCTA, the genetic contribution to the phenotypic correlation was 0.25 (0.11), which is the GCTA genetic correlation weighted by heritability (that is, the product of the square roots of the GCTA heritabilities of g at the two ages). Another way of expressing this is as *bivariate heritability*, which is the proportion of the phenotypic correlation that can be attributed to genetic covariance. GCTA bivariate heritability was 0.60 (that is, $0.25 \div 0.42$), indicating that 60% of the phenotypic correlation could be accounted for by genetic factors. The comparable twin-study estimate of the genetic contribution to the phenotypic correlation was 0.31 (0.03), yielding a bivariate heritability of 0.68.

Increasing heritability

Despite the substantial genetic correlation of 0.73 from age 7–12, GCTA estimates of genetic influence on g increased from 0.26

Table 1. Bivariate GCTA results (with standard errors) for general cognitive ability (g) from age 7 to 12[a]

(a) *Genetic*			
Genetic variance at age 7	*Genetic variance at age 12*	*Genetic covariance between age 7 and 12*	*Genetic correlation between age 7 and 12*
0.26 (.17)	0.45 (0.14)	0.25 (0.11)	0.73 (0.29)
(b) *Environmental*			
Environmental variance at age 7	*Environmental variance at age 12*	*Environmental covariance between age 7 and 12*	*Environmental correlation between age 7 and 12*
0.74 (0.17)	0.55 (0.14)	0.18 (0.11)	0.28 (0.15)

Abbreviation: GCTA, genome-wide complex trait analysis.
[a]GCTA incorporates full-information maximum likelihood that uses the full sample of 2875 individuals with data at either 7 or 12. However, the variance estimates at each age are based on individuals with data at that age (1908 at 7, 2329 at 12) and the covariance estimates are based on individuals with data at both ages (1344).
[b]The current version of GCTA does not report the environmental correlation or its standard error. The environmental correlation was derived here from the GCTA estimates using the following algorithm: C(e)_tr12 / ($\sqrt{}$V(e)_tr1 *$\sqrt{}$V(e)_tr2), whereas the standard error was calculated using: Var(re) = re * re * (VarVe1/(4*Ve1*Ve1) + VarVe2/(4*Ve2*Ve2) + VarCe/(Ce*Ce) + CovVe1Ve2/(2*Ve1*Ve2) - CovVe1Ce/(Ve1*Ce) - CovVe2Ce/(Ve2*Ce)); SE(re) = $\sqrt{}$[Var(re)] where re is the environmental correlation, Ve1 is the residual variance for trait 1, Ce is the residual covariance between two traits, VarVe1 is the sampling variance for Ve1 (residual variance for trait 1), VarCe is the sampling variance for Ce, CovVe1Ve2 is the sampling covariance between Ve1 and Ve2, and CovVe1Ce is the sampling covariance between Ve1 and Ce.

Table 2. Bivariate twin model-fitting results (with standard errors) for general cognitive ability from age 7 to 12[a]

Genetic			
Genetic variance at age 7	*Genetic variance at age 12*	*Genetic covariance between age 7 and 12*	*Genetic correlation between age 7 and 12*
0.36 (0.03)	0.49 (0.04)	0.31 (0.03)	0.75 (0.08)
Shared environment (C)			
C variance at age 7	*C variance at age 12*	*C covariance between age 7 and 12*	*C correlation between age 7 and 12*
0.31 (0.03)	0.19 (0.03)	0.12 (0.03)	0.48 (0.11)
Non-shared environment (E)			
E variance at age 7	*E variance at age 12*	*E covariance between age 7 and 12*	*E correlation between age 7 and 12*
0.33 (0.01)	0.32 (0.01)	0.03 (0.01)	0.09 (0.03)

[a]OpenMx twin model-fitting incorporates full-information maximum likelihood that uses the full sample of 6702 pairs of twins with data at either 7 or 12. However, the variance estimates at each age are based on twin pairs with data at that age (5320 at 7, 4061 at 12), and the covariance estimates are based on twin pairs with data at both ages (2269).

(0.17 s.e.) at age 7 to 0.45 (0.14 s.e.) at age 12, although the large standard errors indicate that the increase did not reach statistical significance. Heritability increased significantly in the twin model-fitting analyses, from 0.36 (0.03) at age 7 to 0.49 (0.03) at age 12. Thus, GCTA estimates account for 74% of the twin-study heritability estimate of g at age 7 and 94% at age 12.

Why genetic stability but increasing heritability?

In summary, GCTA confirms the twin-study hypotheses of strong genetic stability and increasing heritability. In other words, the same genes are largely (about 75%) responsible for genetic influence on g at age 7 and age 12, yet the effect of these genes (heritability) increases substantially from age 7 to 12. How is this possible? We hypothesize that the same genes affect g from age to age but heritability increases as children select their own environments that are correlated with their g-related genetic propensities,[10] a process called genotype – environment correlation.[11] This hypothesis makes three predictions. The first prediction is that g-related experiences will themselves show genetic influence, for which there is considerable evidence from twin studies.[43,44] Second, the links between these experiences and g are expected to be mediated genetically, evidence which is beginning to emerge from twin studies.[45] The third prediction is that genetic links between g and experience should strengthen during development, but this has not yet been investigated. These genetic links are expected especially for experiences in which children are able to select or modify their environments in line with their genetic propensities, in contrast to environments that are passively imposed on children. Supportive evidence to date for this genotype – environment hypothesis relies on twin data, but GCTA can also be used to address these issues with DNA alone.

Genetic architecture

Our GCTA results clarify the genetic architecture of g in ways that are relevant to solving the 'missing heritability' puzzle that has emerged from the limited success of genome-wide association studies to identify the genes responsible for heritability.[46] Two of the major hypotheses to account for missing heritability are epistatic (nonadditive) genetic effects and rare variants, because genome-wide association research is limited to detecting additive genetic effects and genetic effects that can be tagged by the common SNPs used to date on commercially available DNA arrays.[19] Because GCTA is also limited in these same two ways, finding significant GCTA estimates of genetic influence provides strong evidence that current genome-wide association research

strategies can detect the majority of the missing heritability if samples are sufficiently large to provide power to detect associations of small effect size. As noted above, our GCTA estimates of genetic influence account for 74–94% of our twin-study heritability estimates, which implies that most of the missing heritability can be found with additive effects of common SNPs. The heritability that remains missing might be due to epistatic effects and rare variants.

In our longitudinal genetic analyses from age 7 to 12, the GCTA estimate of genetic covariance is also somewhat lower than the twin-study estimate. As shown in Table 1, the genetic covariance for g between ages 7 and 12—that is, the genetic contribution to the phenotypic covariance—is 20% lower for GCTA than for twins (that is, 0.25 for GCTA and 0.31 for twins). However, the GCTA genetic correlation of 0.73 is highly similar to the twin-study genetic correlation of 0.76. The likely reason is that GCTA genetic variance and covariance estimates are attenuated by imperfect linkage disequilibrium between causal variants and genotyped SNPs, but the GCTA estimate of the genetic correlation is unbiased, because the genetic correlation is derived from the ratio between genetic covariance and genetic variance. Because GCTA genetic variance and covariance estimates are biased to the same extent due to imperfect linkage disequilibrium, they cancel each other out in the calculation of the genetic correlation, leaving an unbiased estimate of the genetic correlation.

Implications for brain structure and function

To the extent that g indexes general brain function, the present results suggest hypotheses for the etiology of individual differences in brain development. The same genes can be expected to be responsible for individual differences throughout brain development despite the major mean changes that occur during development. The hypothesis of increasing heritability for individual differences in brain development points to genotype – environment correlation as the process by which genotypes become phenotypes. Importantly, the correspondence between GCTA and twin results indicates that special samples such as twins are no longer needed to test such genetic hypotheses in neurodevelopment—GCTA makes it possible to test them in any large sample of unrelated individuals.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

TEDS is supported by a program grant to RP from the UK Medical Research Council (G0901245, and previously G0500079), with additional support from the US National Institutes of Health (HD044454; HD046167). Genome-wide genotyping was made possible by a grant from the Wellcome Trust to the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z; 085475/Z/08/Z). RP is supported by a Medical Research Council Research Professorship award (G19/2) and a European Advanced Investigator award (295366); MT is supported by a Medical Research Council studentship.

REFERENCES

1 Blakemore SJ. Imaging brain development: the adolescent brain. *Neuroimage* 2012; **61**: 397–406.
2 Power JD, Fair DA, Schlaggar BL, Petersen SE. The development of human functional brain networks. *Neuron* 2010; **67**: 735–748.
3 Giedd JN, Rapoport JL. Structural MRI of pediatric brain development: what have we learned and where are we going? *Neuron* 2010; **67**: 728–734.
4 Plomin R, Kosslyn S. Genes brain and cognition. *Nat Neurosci* 2001; **4**: 1153–1155.
5 Spearman C. 'General intelligence', objectively determined and measured. *Am J Psychol* 1904; **15**: 201–292.
6 Deary IJ. Intelligence. *Ann Rev Psychol* 2012; **63**: 453–482.
7 Deary IJ, Johnson W, Houlihan LM. Genetic foundations of human intelligence. *Hum Genet* 2009; **126**: 215–232.
8 Deary IJ, Penke L, Johnson W. The neuroscience of human intelligence differences. *Nat Rev Neurosci* 2010; **11**: 201–211.
9 Karama S, Colom R, Johnson W, Deary IJ, Haier R, Waber DP *et al.* Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6–18. *Neuroimage* 2011; **55**: 1443–1453.
10 Haworth CMA, Wright MJ, Luciano M, Martin NG, de Geus EJC, van Beijsterveldt CEM *et al.* The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol Psychiatry* 2010; **15**: 1112–1120.
11 Plomin R, DeFries JC, Knopik VS, Neiderhiser JM. *Behavioral Genetics*. 6th edn. Worth: New York, 2013.
12 van Soelen ILC, Brouwer RM, van Leeuwen M, Kahn RS, Pol HEH, Boomsma DI. Heritability of verbal and performance intelligence in a pediatric longitudinal sample. *Twin Res Hum Genet* 2011; **14**: 119–128.
13 Haworth CMA, Plomin R. Quantitative genetics in the era of molecular genetics: learning abilities and disabilities as an example. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 783–793.
14 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–369.
15 Chabris CF, Hebert BM, Benjamin DJ, Beauchamp JP, Cesarini D, van der Loos M *et al.* Most reported genetic associations with general intelligence are probably false positives. *Psychol Sci* 2012; **23**: 1314–1323.
16 Maher B. Personal genomes: the case of the missing heritability. *Nature* 2008; **456**: 18–21.
17 Plomin R. Child development and molecular genetics: 14 years later. *Child Dev* 2012; **30**: 1467–8624.
18 Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76–82.
19 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**: 565–569.
20 Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011; **43**: 519–525.
21 Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 2011; **88**: 294–305.
22 Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 2012; **44**: 247–250.
23 Lubke GH, Hottenga JJ, Walters R, Laurin C, de Geus EJ, Willemsen G *et al.* Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biol Psychiatry* 2012; **72**: 707–709.
24 Vinkhuyzen AAE, Pedersen NL, Yang J, Lee SH, Magnusson PKE, Iacono WG *et al.* Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Transl Psychiatry* 2012; **2**: e102.
25 Benjamin DJ, Cesarini D, van der Loos MJHM, Dawes CT, Koellinger PD, Magnusson PKE *et al.* The genetic architecture of economic and political preferences. *Proc Natl Acad Sci USA* 2012; **109**: 8026–8031.
26 Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D *et al.* Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol Psychiatry* 2011; **16**: 996–1005.
27 Deary IJ, Yang J, Davies G, Harris SE, Tenesa A, Liewald D *et al.* Genetic contributions to stability and change in intelligence from childhood to old age. *Nature* 2012; **482**: 212–215.
28 Plomin R, Haworth CMA, Meaburn EL, Price T. Wellcome Trust Case Control Consortium 2, Davis OSP. Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychol Sci* 2001; **178**: 41–48.
29 Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 2012; **28**: 2540–2542.
30 Haworth CMA, Davis OSP, Plomin R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioural development from childhood to young adulthood. *Twin Res Hum Genet* 2012; **FirstView**: 1–9.
31 Oliver BR, Plomin R. Twins Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behaviour problems from childhood through adolescence. *Twin Res Hum Genet* 2007; **10**: 96–105.
32 Kovas Y, Haworth CMA, Dale PS, Plomin R. The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monogr Soc Res Child Dev* 2007; **72**: 1–144.
33 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.

34 Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007; **317**: 944–947.
35 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; **2**: 2074–2093.
36 Petrill SA, Rempell J, Oliver B, Plomin R. Testing cognitive abilities by telephone in a sample of 6-to 8-year olds. *Intelligence* 2002; **30**: 353–360.
37 Haworth CMA, Harlaar N, Kovas Y, Davis OS, Oliver BR, Hayiou-Thomas ME *et al.* Internet cognitive testing of large samples needed in genetic research. *Twin Res Hum Genet* 2007; **10**: 554–563.
38 van der Waerden BL. On the sources of my book "Moderne Algebra". *Hist Math* 1975; **2**: 31–40.
39 Lehmann EL. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day: San Francisco, 1975.
40 Team RDC. R: A language and environment for statistical computing (Version 2.13.1). 2011, Available from http://www.R-project.org/.
41 Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T *et al.* OpenMx: an open source extended structural equation modeling framework. *Psychometrika* 2011; **76**: 306–317.
42 Davis O, Arden R, Plomin R. g in middle childhood: moderate genetic and shared environmental influence using diverse measures of general cognitive ability at 7, 9 and 10 years in a large population sample of twins. *Intelligence* 2008; **36**: 68–80.
43 Hanscombe KB, Haworth CMA, Davis OSP, Jaffee SR, Plomin R. The nature (and nurture) of children's perceptions of family chaos. *Learn Individ Differ* 2010; **20**: 549–553.
44 Kendler KS, Baker JH. Genetic influences on measures of the environment: a systematic review. *Psychol Med* 2007; **37**: 615–626.
45 Hanscombe KB, Haworth CMA, Davis OSP, Jaffee SR, Plomin R. Chaotic homes and school achievement: a twin study. *J Child Psychol Psychiatry* 2011; **52**: 1212–1220.
46 Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90**: 7–24.

CHAPTER 7: DNA EVIDENCE FOR STRONG GENOME-WIDE PLEIOTROPY OF COGNITIVE AND LEARNING ABILITIES

This chapter is presented as a published paper and is an exact copy of the following journal publication:

Maciej Trzaskowski, Oliver S. P. Davis, John C. DeFries, Jian Yang, Peter M. Visscher, Robert Plomin. (in press). *Behavior Genetics*

Behav Genet
DOI 10.1007/s10519-013-9594-x

ORIGINAL RESEARCH

DNA Evidence for Strong Genome-Wide Pleiotropy of Cognitive and Learning Abilities

Maciej Trzaskowski · Oliver S. P. Davis · John C. DeFries · Jian Yang · Peter M. Visscher · Robert Plomin

Received: 13 February 2013 / Accepted: 10 April 2013

Abstract Very different neurocognitive processes appear to be involved in cognitive abilities such as verbal and non-verbal ability as compared to learning abilities taught in schools such as reading and mathematics. However, twin studies that compare similarity for monozygotic and dizygotic twins suggest that the same genes are largely responsible for genetic influence on these diverse aspects of cognitive function. It is now possible to test this evidence for strong pleiotropy using DNA alone from samples of unrelated individuals. Here we used this new method with 1.7 million DNA markers for a sample of 2,500 unrelated children at age 12 to investigate for the first time the extent of pleiotropy between general cognitive ability (aka intelligence) and learning abilities (reading, mathematics and language skills). We also compared these DNA results to results from twin analyses using the same sample and measures. The DNA-based method revealed strong genome-wide pleiotropy: Genetic correlations were greater than 0.70 between general cognitive ability and language, reading, and mathematics, results that were highly similar to twin study estimates of genetic correlations. These results indicate that genes related to diverse neurocognitive processes have general rather than specific effects.

Keywords Pleiotropy · Intelligence · Learning abilities · Mathematics · Language · GCTA · Twins · Heritability · Cognition

Edited by John K Hewitt.

M. Trzaskowski (✉) · O. S. P. Davis · R. Plomin
MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, DeCrespigny Park, Denmark Hill, London SE5 8AF, UK
e-mail: maciej.trzaskowski@kcl.ac.uk

J. C. DeFries
University of Colorado, Institute for Behavioral Genetics, Boulder, CO 80309, USA

J. Yang · P. M. Visscher
University of Queensland Diamantina Institute, Princess Alexandra Hospital, The University of Queensland, Brisbane, QLD 4102, Australia

P. M. Visscher
Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia

Introduction

Very different neurocognitive processes appear to be involved in cognitive abilities such as reasoning and mathematics (Deary 2000) However, quantitative genetic research, largely based on twin studies, consistently indicates that genes that affect individual differences in performance in one domain are largely the same genes that affect performance in other domains, leading to the Generalist Genes Hypothesis (Plomin and Kovas 2005).

It is now possible to use DNA itself to estimate genetic influence in any sample of unrelated individuals rather than relying on comparisons between monozygotic and dizygotic twins. The method, implemented in a tool called Genome-wide Complex Trait Analysis (GCTA; Yang et al. 2011a) does not attempt to identify specific genes associated with traits. Instead, it correlates genomic similarity across hundreds of thousands of single nucleotide polymorphisms (SNPs) with phenotypic similarity in a large sample of unrelated individuals (Yang et al. 2010). This population-based approach does not rely on the strong assumptions made in classical twin studies.

Univariate Linear Mixed Model (LMM) implemented in the GCTA package has been used to estimate genetic

Published online: 23 April 2013

influence for height and body mass index (Yang et al. 2010, 2011b), psychiatric and medical disorders (Lee et al. 2011), personality (Vinkhuyzen et al. 2012), and cognitive abilities (Davies et al. 2011; Plomin et al. 2013b). In contrast to univariate genetic analysis, bivariate genetic analysis focuses on the genetic correlation, the correlation between genetic influences on different traits, called pleiotropy (Plomin et al. 2013a). High genetic correlations between phenotypes are often interpreted as an indication that the same genes affect the phenotypes. Genetic correlations between diverse cognitive abilities as estimated through twin studies are typically greater than 0.60, indicating that cognition-related genes largely have general pleiotropic effects (Calvin et al. 2012; Plomin and Kovas 2005). However, the genetic correlation estimated from twin studies could be biased due to misspecification of the model of twin similarity for genetic and non-genetic effects. In this study, we use the GCTA package to estimate the genetic correlation between traits in conventionally unrelated individuals based on DNA evidence alone; this estimate is free of bias if we assume that the sole reason for phenotypic similarity between conventionally unrelated individuals is shared additive genetic factors. For brevity, we refer to LMM used in the GCTA package simply as GCTA.

Here we use bivariate GCTA (Lee et al. 2012; Yang et al. 2011a) to test the Generalist Genes Hypothesis by estimating genetic correlations between general cognitive ability ('g', aka intelligence) and language, reading, and mathematics. We compare these genetic correlation estimates from GCTA to those obtained from the twin design using the same sample assessed at the same age with the same measures. We also analyze the variables of height and weight for purposes of comparison.

Materials and methods

Sample and genotyping

The sample was drawn from the Twins Early Development Study (TEDS), which is a multivariate longitudinal study that recruited over 11,000 twin pairs born in England and Wales in 1994, 1995 and 1996 (Haworth et al. 2012; Oliver and Plomin 2007). TEDS has been shown to be representative of the UK population (Kovas et al. 2007). The project received approval from the Institute of Psychiatry ethics committee (05/Q0706/228) and parental consent was obtained prior to data collection.

Cognitive and DNA data were available for 3,747 11- and 12-year-old children whose first language was English and had no major medical or psychiatric problems. From that sample, 3,665 DNA samples were successfully hybridized to Affymetrix GeneChip 6.0 SNP genotyping arrays using standard experimental protocols as part of the WTCCC2 project (for details see Trzaskowski et al. 2013). In addition to nearly 700,000 genotyped SNPs, more than one million other SNPs were imputed from HapMap 2, 3 and WTCCC controls using IMPUTE v.2 software (Howie et al. 2009). 3,152 DNA samples (1,446 males and 1,706 females) survived quality control criteria for ancestry, heterozygosity, relatedness, and hybridization intensity outliers. To control for ancestral stratification, we performed principal component analyses on a subset of 100,000 quality-controlled SNPs after removing SNPs in linkage disequilibrium ($r^2 > 0.2$) (Fellay et al. 2007). Using the Tracy–Widom test (Patterson et al. 2006), we identified 8 axes with $p < 0.05$, which were used as covariates in GCTA analyses.

The mean age of the sample was 11.5 years (SD = 0.66). The sample sizes for the GCTA results shown in Table 1 are 2,325 for 'g' and language, 2,238 for 'g' and mathematics, 2,250 for 'g' and reading, and 2,296 for height and weight. For the twin analyses, cognitive data were available for 5,434 twin pairs (Davis et al. 2009); however, the twin analyses presented here were based only on twins included in the GCTA analyses in order to provide a more precise comparison between GCTA and twin-study results. The numbers of twin pairs were 2,205, 2,095, 2,104 and 2,162, respectively.

Measures

Cognitive data were collected online via the Internet using, where possible, adaptive branching, which enabled measurement of the full range of ability using a relatively small number of items. Details about the following measures, including references, are available elsewhere (Kovas et al. 2007).

General cognitive ability (g)

'g' was assessed from two verbal tests and two non-verbal tests. The verbal tests included WISC-III-PI Multiple Choice Information (General Knowledge) and Vocabulary

Table 1 Genome-wide Complex Trait Analysis (GCTA) and twin study estimates of genetic correlations. Standard errors (SE) are shown in parentheses. 'g' refers to general cognitive ability

Bivariate comparison	Genetic correlation	
	GCTA (SE)	Twin (SE)
'g' vs language	0.81 (0.15)	0.80 (0.06)
'g' vs mathematics	0.74 (0.15)	0.73 (0.03)
'g' vs reading	0.89 (0.26)	0.66 (0.05)
'g' vs height	−0.13 (0.30)	−0.03 (0.06)
'g' vs weight	−0.04 (0.25)	−0.06 (0.06)
Height vs weight	0.76 (0.13)	0.65 (0.02)

Multiple Choice subtest. The two non-verbal reasoning tests were WISC-III-UK Picture Completion and Raven's Standard and Advanced Progressive Matrices.

Language

Three components of language were assessed: syntax, semantics and pragmatics. Syntax was measured using the Listening Grammar subtest of the Test of Adolescent and Adult Language. Semantics was assessed using Level 2 of the Figurative Language subtest of the Test of Language Competence. Pragmatics was assessed using Level 2 of the Making Inferences subtest of the Test of Language Competence.

Mathematics

Assessment of mathematics targeted three components of mathematics: Understanding Number, Non-numerical Processes, and Computation and Knowledge. The items for these three scales were based on the National Foundation of Educational Research 5–14 Mathematics Series.

Reading

Four measures of reading were employed. Two measures assessed reading comprehension: the reading comprehension subtest of the Peabody Individual Achievement Test and the GOAL Formative Assessment in Literacy for Key Stage 3. Reading fluency was assessed by an adaptation of the Woodcock–Johnson III Reading Fluency Test and by the Test of Word Reading Efficiency, which was administered by telephone.

Composite measures for 'g', language, mathematics, and reading. For each cognitive measure, outliers above or below 3 SD from the mean were excluded. Scores were regressed on sex and age, and standardized residuals were derived and quantile normalized (Lehmann 1975; van der Waerden 1975). Composite measures for 'g', language, mathematics, and reading were created as unit-weighted means requiring complete data for at least 3 of the 4 tests for 'g' and reading and 2 of 3 tests for language and mathematics. All procedures were executed using R *(*www.r-project.org; R Development Core Team 2011*)*. The phenotypic correlations among the composite measures were 0.63 for 'g' and language, 0.63 for 'g' and mathematics, and 0.57 for 'g' and reading.

Height and weight

Height and weight were assessed on the same sample (age 12) via self-report. Similar to the cognitive measures, outliers (± 3SD) were removed and scores were controlled for age and sex. The phenotypic correlation between height and weight was 0.63.

Statistical analyses

GCTA

Conceptually, the amount of phenotypic variance, or covariance, explained by genetic factors is estimated by a comparison of a matrix of pairwise genomic similarity to a matrix of pairwise phenotypic similarity (Yang et al. 2010). Before the variance or covariance can be decomposed into genetic and residual components, we need to calculate pairwise genomic similarity between all pairs of individuals in the sample using all genetic markers genotyped on the SNP array. Because the GCTA package uses a random effects model to estimate genetic effects from a sample of unrelated individuals in the population, any pair whose genetic similarity is equal to or greater than a fourth cousin is removed (estimate of pairwise relatedness >0.025). In univariate analysis, the variance of a trait can be partitioned using residual maximum likelihood into genetic and residual components. Detailed description of this method can be found in GCTA publications (Yang et al. 2010, 2011a, b). The bivariate method extends the univariate model by relating the pairwise genetic similarity matrix to a phenotypic covariance matrix between traits 1 and 2 (Lee et al. 2012). The eight principal components described earlier were used as covariates in our bivariate GCTA analyses; as mentioned in the previous section, all phenotypes were age- and sex-regressed prior to analysis.

Twin modelling. The twin design and model-fitting is discussed elsewhere (Plomin et al. 2013a). We fit a bivariate Cholesky decomposition using OpenMx (Boker et al. 2011), which provided a direct comparison with the bivariate GCTA. The correlated factor solution is the least restricted model allowing variables to correlate with one another via genetic, shared environment, and non-shared environment. Because previous analyses of these data indicated nonsignificant differences in model-fitting results between males and females (Kovas et al. 2007), we combined same-sex and opposite-sex DZ twin pairs in order to increase the power of the analyses.

Results

Table 1 shows GCTA-estimated genetic correlations (and standard errors, SE) between 'g' and learning abilities for more than 2,238 12-year-old UK twins (randomly selecting only one member of each twin pair to control for potential confounds, such as birth order) based on 1.7 million SNPs measured from the Affymetrix 6.0 GeneChip or imputed from HapMap 2,3 and WTCCC controls (Trzaskowski et al. 2013). Genetic correlations are significant and

Table 2 Bivariate Genome-wide Complex Trait Analysis (GCTA) results (with standard errors) for general cognitive ability ('g') versus language, mathematics, and reading, as well as comparison data for: g and height, g and weight, and height and weight

Variables	A						E				Vp_tr1	Vp_tr2	n
	V(G)_tr1	V(G)_tr2	C(G)_tr12	V(G)/Vp_tr1	V(G)/Vp_tr2	r_G	V(e)_tr1	V(e)_tr2	C(e)_tr12	r_E^a			
'g' vs language	0.36(0.14)	0.35(0.14)	0.29(0.12)	0.37(0.14)	0.35(0.14)	0.81(0.15)	0.63(0.14)	0.65(0.14)	0.33(0.12)	0.52(0.11)	0.99(0.03)	1(0.03)	2325
'g' vs maths	0.36(0.13)	0.32(0.12)	0.25(0.10)	0.36(0.13)	0.32(0.12)	0.74(0.15)	0.64(0.13)	0.67(0.12)	0.38(0.10)	0.57(0.09)	0.99(0.03)	1(0.03)	2238
'g' vs reading	0.34(0.13)	0.16(0.12)	0.20(0.10)	0.34(0.13)	0.16(0.12)	0.89(0.26)	0.65(0.13)	0.84(0.12)	0.36(0.10)	0.49(0.09)	0.99(0.03)	1(0.03)	2259
'g' vs height	0.34(0.14)	0.36(0.15)	−0.05(0.10)	0.35(0.14)	0.36(0.14)	−0.13(0.30)	0.65(0.14)	0.64(0.14)	0.11(0.10)	0.17(0.16)	0.99(0.03)	1(0.03)	1868
'g' vs weight	0.35(0.14)	0.47(0.15)	−0.02(0.10)	0.35(0.14)	0.47(0.14)	−0.04(0.25)	0.64(0.14)	0.53(0.14)	0.04(0.10)	0.07(0.17)	0.99(0.03)	1(0.03)	1868
height vs weight	0.37(0.15)	0.49(0.15)	0.33(0.12)	0.37(0.14)	0.48(14)	0.76(0.13)	0.63(0.14)	0.52(0.14)	0.32(0.12)	0.56(0.12)	1.0(0.03)	1(0.03)	2286

GCTA incorporates full-information maximum likelihood that uses the full sample of more than 2,900 individuals with data on trait 1 or trait 2. However, the variance estimates for each trait are based on individuals with data for that trait and the covariance estimates are based on individuals with data for both traits. The n reported in the last column is the most conservative, i.e., the n that was used for the estimation of the covariance *V(G)* variance explained by genetic factors for trait 1 and trait 2 (tr1, tr2), *C(G)* covariance between trait 1 and 2 explained by genetic factors; *V(e)* residual variance for trait 1 and trait 2, *C(e)* residual covariance between trait 1 and trait 2; *Vp* phenotypic variance for trait 1 and trait 2, *V(G)/Vp* proportion of the phenotypic variance explained by genetic factors for trait 1 and trait 2, r_G genetic correlation between trait 1 and trait 2; *logL* log likelihood estimation of the model, *n* number of individuals with data for both trait 1 and trait 2, values in parentheses are standard errors

[a] The current version of GCTA does not report the environmental correlation or its standard error. The environmental correlation was derived here from the GCTA estimates using the following algorithm: C(e)_tr12/($\sqrt{}$V(e)_tr1 × $\sqrt{}$V(e)_tr2), whereas the standard error was calculated using: Var(re) = re × re × (VarVe1/(4 × Ve1 × Ve1) + VarVe2/(4 × Ve2 × EV) + VarCe/(Ce × Ce) + CovVe1Ve2/(2 × Ve1 × Ve2 − CovVe1Ce/(Ve1 × Ce) − CovVe2Ce/(Ve2 × Ce)); SE(re) = sqrt[Var(re)], where re is the environmental correlation, Ve1 is the residual variance for trait 1, Ce is the residual covariance between two traits, VarVe1 is the sampling variance for Ve1 (residual variance for trait 1), VarCe is the sampling variance for Ce, CovVe1Ve2 is the sampling covariance between Ve1 and Ve2, and CovVe1Ce is the sampling covariance between Ve1 and Ce

Table 3 Bivariate twin model-fitting results (with standard errors) for general cognitive ability ('g') versus language, mathematics, and reading, as well as comparison data for: g and height, g and weight, and height and weight

Variables	A				C				E				n/pairs
	V(G)_tr1	V(G)_tr2	C(G)_tr12	r_G	V(c)_tr1	V(c)_tr2	C(c)_tr12	r_C	V(e)_tr1	V(e)_tr2	C(e)_tr12	r_E	
'g' vs language	0.47(0.05)	0.41(0.05)	0.36(0.04)	0.80(0.06)	0.21(0.05)	0.22(0.04)	0.19(0.03)	0.90(0.10)	0.33(0.02)	0.37(0.02)	0.09(0.01)	0.27(0.03)	2205
'g' vs maths	0.46(0.05)	0.48(0.04)	0.36(0.03)	0.73(0.03)	0.21(0.04)	0.20(0.04)	0.19(0.03)	1.0(0.10)	0.33(0.02)	0.32(0.02)	0.07(0.01)	0.23(0.03)	2095
'g' vs reading	0.46(0.05)	0.59(0.04)	0.34(0.03)	0.66(0.05)	0.21(0.04)	0.17(0.04)	0.15(0.03)	0.85(0.12)	0.33(0.02)	0.24(0.01)	0.06(0.01)	0.20(0.04)	2104
'g' vs height	0.48(0.05)	0.80(0.04)	−0.02(0.03)	−0.03(0.06)	0.19(0.04)	0.10(0.04)	0.08(0.03)	0.54(0.23)	0.33(0.02)	0.10(0.01)	0.01(0.01)	0.07(0.04)	1716
'g' vs weight	0.48(0.05)	0.83(0.03)	−0.04(0.04)	−0.06(0.06)	0.19(0.04)	0.05(0.03)	0.03(0.03)	0.32(0.51)	0.33(0.02)	0.12(0.01)	0.03(0.01)	0.13(0.04)	1716
Height vs weight	0.81(0.04)	0.85(0.04)	0.54(0.03)	0.65(0.02)	0.09(0.04)	0.04(0.02)	0.06(0.03)	1.0(0.00)	0.10(0.01)	0.11(0.01)	0.06(0.01)	0.41(0.03)	2162

OpenMx twin model-fitting incorporates full-information maximum likelihood that uses the full sample of more than 2,000 pairs of twins with data on trait 1 or trait 2. However, the variance estimates for each trait are based on individuals with data for that trait. The covariance estimates are based on twin pairs with data for both traits, which is the conservative sample size shown in the last column

V(G) proportion of the variance explained by genetic factors for trait 1 and trait 2 (tr1, tr2), *C(G)* proportion of the covariance between trait 1 and 2 explained by genetic factors; *V(c)* proportion of the variance explained by shared environment for trait 1 and trait 2 (tr1, tr2), *C(c)* proportion of the covariance between trait 1 and 2 explained by shared environment, *V(e)* proportion of the variance explained by non-shared environment for trait 1 and trait 2 (tr1, tr2), *C(e)* proportion of the covariance between trait 1 and 2 explained by non-shared environment, *rG* genetic correlation; *rC* correlation of shared environmental factors; *rE* correlation of non-shared environmental factors, *n* number of twin pairs with data for both trait 1 and trait 2; values in parentheses are standard errors

substantial for all three comparisons—between 'g' and language (0.81), mathematics (0.74), and reading (0.89). The GCTA-estimated genetic correlations between 'g' and learning abilities are similar in magnitude to the GCTA-estimated genetic correlation between height and weight (0.76). In addition, Table 1 includes bivariate results for 'g' versus height and 'g' versus weight as 'negative controls'; their phenotypic correlations are both 0.07. As expected, these comparisons yielded negligible and nonsignificant genetic correlations (−0.03 and −0.06, respectively).

Table 1 also includes analogous genetic correlations from twin model-fitting analyses, as estimated from the same twin sample but including the co-twins (more than 2,095 pairs of twins). The GCTA-estimated genetic correlations are highly similar to the twin study estimates and do not differ significantly, as indicated by their overlapping standard errors. The similarity of GCTA and twin estimates of genetic correlations extend to the comparison between height and weight as well as the negative control comparisons of 'g' and height and 'g' and weight.

Tables 2 and 3 show full results from the bivariate GCTA and twin analyses, respectively.

Discussion

Using DNA evidence alone, these high genetic correlations estimated from GCTA support the Generalist Genes Hypothesis in showing strong pleiotropy between 'g' and learning abilities, especially because we show that these GCTA-estimated genetic correlations are as high as genetic correlations estimated from the twin design.

Although GCTA does not identify specific genes associated with these traits, it addresses a critical issue in genome-wide association studies: the extent to which common SNPs used on commercially available SNP arrays can account for the heritability of quantitative traits (Yang et al. 2011b). We have shown in univariate GCTA analyses that, if samples were sufficiently large, common SNPs could account for more than two-thirds of the heritability of cognitive abilities estimated in twin studies (Yang et al. 2011b; see also Table 2). Why are univariate GCTA heritability estimates less than the twin study estimates of heritability? As discussed elsewhere (e.g. Yang et al. 2010), the main problem is imperfect tagging. The common SNPs used on all available commercial arrays only capture what is in LD with them. Rare variants, which have lower minor allele frequency, will thus not be 'tagged' and their influence will be missed. In addition, GCTA estimates additive genetic influence only, so that non-additive effects (gene–gene and gene-environment interaction) are not captured either.

A more novel question, and central to the present paper, is why, as we have shown here, bivariate genetic

correlations estimated by GCTA are as great as twin study estimates. The likely reason is that attenuation of the estimated additive genetic variance due to imperfect linkage disequilibrium between causal variants and genotyped SNPs applies to both the additive genetic variance of the two traits and to their additive genetic covariance by the same proportion. Thus, the GCTA estimate of the genetic correlation is unbiased because it is derived from the ratio between genetic covariance and the genetic variances of the two traits.

Are generalist genes all in the mind (cognition) or are they in the brain as well? That is, genetic correlations between cognitive and learning abilities might be epiphenomenal in the sense that multiple genetically independent brain mechanisms could affect each ability, creating genetic correlations among abilities. However, the genetic principles of pleiotropy (each gene affects many traits) and polygenicity (many genes affect each trait) lead us to predict that generalist genes have their effects further upstream, creating genetic correlations among brain structures and functions, a prediction that supports a network view of brain structure and function.

Acknowledgments Twins Early Development Study (TEDS) is supported by a program grant to RP from the UK Medical Research Council (G0901245, and previously G0500079), with additional support from the US National Institutes of Health (HD044454; HD046167). Genome-wide genotyping was made possible by a grant from the Wellcome Trust to the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z; 085475/Z/08/Z). RP is supported by a Medical Research Council Research Professorship award (G19/2) and a European Advanced Investigator award (295366); MT is supported by a Medical Research Council studentship; OSPD is supported by a Sir Henry Wellcome Fellowship from the Wellcome Trust (WT088984). The authors declare that they had no conflicts of interest with respect to their authorship or publication of this article.

References

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J (2011) OpenMx: an open source extended structural equation modeling framework. Psychometrika 76(2):306–317

Calvin C, Deary I, Webbink D, Smith P, Fernandes C, Lee S, Luciano M, Visscher P (2012) Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. Behav Genet 42(5):699–710

Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, Ke X, Le Hellard S, Christoforou A, Luciano M, McGhee K, Lopez L, Gow AJ, Corley J, Redmond P, Fox HC, Haggarty P, Whalley LJ, McNeill G, Goddard ME, Espeseth T, Lundervold AJ, Reinvang I, Pickles A, Steen VM, Ollier W, Porteous DJ, Horan M, Starr JM, Pendleton N, Visscher PM, Deary IJ (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 16(10):996–1005

Davis OSP, Haworth CMA, Plomin R (2009) Learning abilities and disabilities: generalist genes in early adolescence. Cog Neuropsychiatry 14(4–5):14312–14331

Deary I (2000) Looking down on human intelligence: from psychometrics to the brain. Oxford University Press, Oxford

Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317(5840):944–947

Haworth CMA, Davis OSP, Plomin R (2012) Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioural development from childhood to young adulthood. Twin Res Hum Genet 16(1):117–125

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5(6):e1000529

Kovas Y, Haworth CMA, Dale PS, Plomin R (2007) The genetic and environmental origins of learning abilities and disabilities in the early school years. Monogrs Soc Res Child Dev 72(3):144–721

Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet 88(3):294–305

Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR (2012) Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. Bioinformatics 28(19):2540–2542

Lehmann EL (1975) Nonparametrics: statistical methods based on ranks. Holden-Day, San Francisco

Oliver BR, Plomin R (2007) Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems from childhood through adolescence. Twin Res Human Genet 10(1):96–105

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12):2074–2093

Plomin R, Kovas Y (2005) Generalist genes and learning disabilities. Psychol Bull 131:592–617

Plomin R, DeFries JC, Knopik VS, Neiderhiser JM (2013a) Behavioral genetics. Worth Publishers, New York

Plomin R, Haworth CMA, Meaburn EL, Price TS, Wellcome Trust Case Control Consortium 2, Davis OSP (2013b) Common DNA markers can account for more than half of the genetic influence on cognitive abilities. Psychol Sci 24(4):562–568

R Development Core Team (2011) R: a Language and environment for statistical computing. Vienna, Austria

Trzaskowski M, Eley TC, Davis OSP, Doherty SJ, Hanscombe KB, Meaburn EL, Haworth CMA, Price T, Plomin R (2013) First genome-wide association study on anxiety-related behaviours in childhood. PLoS One 8(4):e58676

van der Waerden BL (1975) On the sources of my book "Moderne Algebra". Historia Mathematica 2:31–40

Vinkhuyzen AA, Pedersen NL, Yang J, Lee SH, Magnusson PK, Iacono WG, McGue M, Madden PA, Heath AC, Luciano M, Payton A, Horan M, Ollier W, Pendleton N, Deary IJ, Montgomery GW, Martin NG, Visscher PM, Wray NR (2012) Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. Transl Psychiatry 2:e102

Yang J, Benyamin B, McEvoy BP, Gordon SD, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42(7):565–569

Yang JA, Lee SH, Goddard ME, Visscher PM (2011a) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1):76–82

Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, De Andrade M, Feenstra B, Feingold E, Hayes MG (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 43(6):519–525

CHAPTER 8: DISCUSSION

The three key aims of this thesis were: (1) to compare the total amount of variance explained by a newly developed molecular method called Genome-wide Complex Trait Analysis (GCTA) with twin-estimated heritability in the same sample of individuals to quantify the extent of the 'missing heritability' (Chapters 2, 3 and 4); (2) to compare twin and GCTA data for increasing heritability across development in the presence of strong genetic stability (Chapters 5 and 6); (3) to compare twin and GCTA data for genetic pleiotropy (Chapter 7). I summarise the studies described in this thesis, then discuss limitations, and conclude by exploring implications and future directions.

SUMMARY OF FINDINGS

In Chapter 2 I investigated the extent of the 'missing heritability' in a set of anxiety-related traits by performing, for the first time, genome-wide association (GWA) and Genome-wide Complex Trait Analysis (GCTA). I showed that common variants of large effect size do not explain individual differences in anxiety-related traits (as found with other complex traits). Instead, GWA results indicate that these complex traits are either driven by numerous common single nucleotide polymorphisms (SNPs) of very small effect size, by rare variants of larger effect size, by non-additive genetic effects, or by a combination of all of these. To probe these explanations further, I conducted GCTA on the same phenotypes. The results of those analyses suggest that, in anxiety-related traits, the additive effects of common SNPs play a minor role.

This small influence of common SNPs on anxiety-related traits was surprising, especially since the study was adequately powered to detect moderate GCTA heritability. One outstanding question is why the estimates of genetic influence on individual differences in anxiety-related traits differs between alternative genetic methods such as fitting twin data to structural equation models versus GCTA which is another kind of quantitative genetic analysis. One possibility is that the gap between the two estimates is caused by rare SNPs

of larger effect size that are not assessed by GCTA (but are captured in twin model-fitting analyses). Another possibility is that non-additive genetic effects contribute to the gap in estimates for the same reason – they are not observable in GCTA. There is a third possibility worth considering: rater bias. In our study the anxiety-related traits consisted of an observation made by one parent. In GCTA, parents rated only one child, whereas in the twin model each parent rated two siblings. If parents of MZ twins rated them more similarly than did parents of DZ twins, the estimate of anxiety heritability derived from model-fitting would be inflated. If the twin heritability was inflated by rater bias and its true estimate was less than the reported 50-60% then our power to detect the genetic influence using GCTA would be inadequate. To test this, I ran another experiment described in Chapter 3.

In Chapter 3 I asked whether the low genetic estimates of the first GCTA study on anxiety were confined to parent ratings or if they generalise to other raters and other behaviour problems. In order to answer this, I conducted univariate twin model-fitting as well as GCTA analyses of several behaviour problems in childhood that were rated by parents, teachers and children themselves. The results of the univariate twin analyses showed that behaviour problems are moderately to highly heritable. Our GCTA results suggested that, regardless of the rater or the particular behaviour problem, the additive effect of common SNPs is small. Together these results showed that the gap in estimates of genetic influence between classical twin model-fitting and GCTA was not a special property of anxiety-related traits but rather a trend general to most behaviour problems.

To rule out the possibility that our sample was different from other reported samples, we compared our analyses of behaviour problems with analyses of cognitive and anthropometric phenotypes. The GCTA estimates of genetic influence on these measures were within the expected ~50% of twin heritability; this militates strongly against the likelihood that our sample is unrepresentative. To what should we attribute the difference in estimate of genetic influence? One possibility is that non-additive effects and rare variants exert significant influences on the phenotypes but are not yet modelled in GCTA because rare variants are not yet reliably assessed by microarrays and non-additive effects are not captured by GCTA. In propose that we should

discriminate between two kinds of 'missing heritability' that may have different solutions. We should investigate 'GWA missing heritability' and 'GCTA missing heritability'. The former refers to the proportion of additive effects of common SNPs that may yet be discovered; the latter refers to rare variants and non-additive genetic influences, not yet assayed.

In a final study investigating missing heritability, I explored body mass index (BMI); a phenotype for which GWA studies have had some success. Twin study estimates of genetic influence on BMI from classical structural equation model-fitting are typically around 50-90% (Elks et al, 2012). This differs sharply from the 2% of variation accounted for by the 32 SNPS (in aggregate) identified by GWA studies. In addition, GCTA studies of adults have found that the additive influence of common SNPs accounts for approximately 15% of the variance in BMI or ~30% of the heritability estimate derived from classical twin model-fitting. However these findings are limited to adults, and no direct comparison between twin and GCTA analyses have been made. To extend what is known about the aetiology BMI, I conducted studies on a large community sample of children with three analytical methods: classical twin model-fitting, GWA and GCTA. The univariate twin model-fitting and GCTA analyses showed that the additive effect of common SNPs could explain nearly 40% of twin heritability of BMI at age 10. Although the amount of variance explained by common SNPs in childhood was only markedly larger than that explained in adulthood, I argued that this was plausible because adults may be more likely than children to be making weight loss attempts, and so taking active steps towards controlling their environment.

In a previous publication we have shown that twin heritability of BMI increases across childhood (Haworth et al., 2008), but this has not been confirmed using DNA alone. In Chapter 5 I therefore extended the investigation into BMI to longitudinal analyses to test whether the increasing heritability between age 4 and age 10 can be replicated using DNA alone.

Making sense of the phenomenon of increasing heritability through the life course using GCTA is useful partly because the phenomenon itself is not perturbed by the fact that GCTA may under-estimate heritability

(because it is limited to the variance captured by commercial arrays). That is, the extent to which GCTA under-estimates heritability should not differ whether applied to age 4 or 10. Another way to investigate change across time is to examine longitudinal genetic correlations. Previous longitudinal twin studies on childhood BMI showed that heritability increases despite strong genetic correlation. The genetic correlation is calculated from estimates of the phenotypic variance covariance matrix among twins. The phenotypic correlation is a covariance divided by a product of two standard deviations (Equation 4). The proportion of the covariance due to genetic factors is a product of the square roots of the two heritabilities and the genetic correlation. Thus, when calculating the genetic correlation, the heritability terms are in both the nominator and the denominator (Equation 5). Consequently, if heritability estimated by GCTA is biased, the same bias is present in the nominator as well as the denominator and cancels out.

Equation 4. Phenotypic correlation

$$r_{xy} = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

Equation 5. Genetic correlation

$$r_{g_{xy}} = \frac{Cov_g(x,y)}{\sqrt{h_x h_y}} = \frac{\sqrt{h_x h_y} r_{g_{xy}}}{\sqrt{h_x h_y}}$$

Examining BMI offered an opportunity to compare directly the results of twin model-fitting to GCTA as well as using a set of polygenic predictors. This methodological triangulation of analyses contributed a novel way to illuminate this complex trait. First, I applied bivariate twin model-fitting techniques and GCTA models to the same sample measured for BMI at ages 4 and 10 to test whether increasing heritability and genetic stability could be replicated using DNA alone. As predicted, the trend of increasing heritability emerged from both methods. In addition, the magnitude of the genetic correlations was almost identical in both the GCTA and the classical twin analysis. This meant three things: these phenomena were 'real', twin results did not inflate

genetic influence, and gene-environment correlation was a strong candidate responsible for this increase. The genetic correlations did not reach unity, so 'new' genes that influence BMI may come 'on stream' between ages 4 and 10. To test more precisely whether the same genes could increase their association with BMI over time, I turned to a set of 32 SNPs that had been associated with BMI through GWA meta-analysis (Speliotes et al., 2010). This enabled me to test whether the same SNPs would show increasing effects as children get older. The results showed a significant increase in variance explained from age 4 to 10, suggesting that the trend of increasing heritability can now be replicated using specific DNA markers. In addition, finding that the same SNPs show increasing associations with BMI from age 4 to 10 increases our confidence that the high genetic correlations reported from both GCTA and twin analyses are 'true' and that gene-environment correlation plays a role in explaining the rise in heritability over development.

As noted earlier GCTA results explained a larger proportion of heritability in children than previously reported in adults. This suggests that common SNPs exert a stronger influence on BMI in childhood. However, I have shown that heritability of BMI also increases during childhood. Although puzzling, this non-linear life-long trend could be a function of environmental control. Specifically, increasing heritability in childhood could be a result of progressive autonomy; parents control the feeding routines of young children limiting the potential for genetic expression through child eating behaviour (age 4), but as children get older and gain independence the heritability of BMI rises (age 10); on the other hand, a decrease in heritability in adulthood could reflect adults making active attempts to control their weight through diet and exercise.

To further substantiate the finding of increasing heritability and genetic stability in BMI I conducted similar analyses in another phenotype that also showed the same longitudinal patterns, intelligence. Although no set of specific SNPs is yet associated with this phenotype, I was able to test the hypothesis by directly comparing bivariate twin and GCTA models at ages 7 and 12 (Chapter 6). As in the longitudinal study of BMI reported in Chapter 5, the magnitude of age-to-age genetic correlations for intelligence was similar in twin and GCTA analyses, supporting the idea that the same genes can exert greater influence on phenotypes across ages.

In Chapter 7, I extended the direct comparison between twin-modelling and GCTA methods to another commonly reported finding: pleiotropy. It has been reported repeatedly that genes tend to be general, i.e. each gene influences many traits. I set out to explore pleiotropy by applying bivariate twin and GCTA models to intelligence and learning abilities and just as in the two previous studies (Chapters 5 and 6), I found that GCTA- and twin-estimated between-traits genetic correlations were of similar magnitude. This finding reinforced the conclusion that what we have learned from twin studies can be replicated using DNA alone. This should reassure us that the classical twin method is useful and reliable for exploring the aetiology and genetic architecture of complex traits. Unlike GCTA, the twin method can tease apart the influences of shared and unique environments.

LIMITATIONS

Many limitations of the studies in this thesis have been discussed within the respective chapters. However, there are several that apply to this work in general. One notable limitation of these studies is that they focus specifically on the genetic contribution to traits, even though environmental contributions are also important. The reason for the emphasis on genetics is that GCTA does not distinguish between non-genetic familial factors and factors unique to each individual. The unexplained residual in a GCTA analysis comprises any influence on the trait that is not the additive effect of common SNPs. These include rare variants, epistasis, gene-environment interplay, unique environment as well as error of measurement. As will be discussed in the 'Implications' section, perhaps the best way to study the environment is through polygenic predictors. Currently most complex traits lack genetic candidates that could be used to create these predictors. It is important to study the impact of the environment but statistical tools vary in their capacity to elucidate different components of influence on traits.

Although I argued in Chapters 5 and 6 that the developmental increase in heritability is driven by gene-environment correlation, I could not rule out other phenomena such as gene-environment interaction. Specifically, the increase in heritability as estimated through the twin method was markedly stronger than the

increase estimated through GCTA, which could indicate rare or non-additive genetic influences coming online later in childhood. These new genetic influences could also increase our susceptibility to the environment (gene-environment interaction) or exert a stronger need to 'seek out' the 'right' environments (gene-environment correlation). What we have learnt about complex traits hitherto would suggest that all of these factors should be of some importance, but the exact nature of the influence is yet to be determined.

Finally, power for GCTA was limited. Although the results from the twin models had small standard errors suggesting we had adequate power to reliably quantify genetic influences, GCTA estimates had larger standard errors. As a result, the comparison between the molecular and twin techniques mainly focused on the consistency of point estimates across time and traits, rather than on the levels of significance.

Implications and Future Directions

I have endeavoured in this thesis to make a substantial contribution to testing competing explanations of 'missing heritability'. I have explored several complex traits through a set of different methods. A novel contribution to the field is the direct comparison of parameter estimates between the classical twin and molecular methods in the same sample of children using the same measures at the same ages. The direct comparison between the twin and GCTA methods at the univariate level has shown that common SNPs play an important role in individual differences in some but not all complex traits. It also highlighted the importance of not-yet-captured rare variants and non-additive effects. At the bivariate level, the comparison provided strong support for the accuracy of the twin results highlighting the usefulness of the twin method for future genetically-sensitive studies.

Despite GWA potentials, the methods used in this thesis could not show which genes influence these complex traits, nor could they show whether rare variants or non-additive influences were implicated. To gain better insights into these issues we will require more detailed profiling of the human genome and larger samples. How else should we continue to make progress in understanding the aetiology of complex phenotypes?

Several avenues of research look promising including next generation sequencing, non coding RNAs and animal models.

NEXT GENERATION SEQUENCING

Next generation sequencing could identify more genetic loci associated with complex traits particularly because it may illuminate the extent to which rare variants matter. These hard to capture variants may play an important role in the extreme tails of a trait's distribution. For example, we found that common SNPs influence obesity across the distribution of the normal population. However, very rare mendelian mutations may account for about 5% of extreme obesity (Farooqi & O'Rahilly, 2006). Another study showed that rare variants – copy number variants (CNVs) – influence severe early onset obesity (Wheeler et al., 2013).

Other clinical phenotypes have been associated with an increased burden of rare variants (Girirajan et al., 2012; Zahnleiter et al., 2013). Perhaps rare variants are responsible for a puzzle in findings concerning mild and severe intellectual disability. One study found that mild intellectual impairment was familial, but severe impairment was not (Nichols, 1984). The researchers found that the siblings of severely mentally retarded children showed no mental impairment, whereas siblings of children with 'mild' mental disability did. Common (less severe) disorders are likely to be the quantitative extremes of normal variation (Plomin, Haworth, & Davis, 2009), whereas extreme levels of disability could be mainly a result of accumulation of much rarer variants or private *de novo* mutations.

Greater characterisation of the genome could result in an increased number of associated DNA variants, which then could be used as polygenic predictors (PGS) to study the environment. Having a set of genetic loci that have been robustly associated with a trait through GWA, enables more precise testing of gene-environment interactions, gene-environment correlations (Qi et al., 2012), and also mediation pathways. For example, in another publication, we have shown that common obesity-risk increasing genes influence adiposity partly via appetitive mechanisms (Llewellyn, Trzaskowski, van Jaarsveld, Plomin, & Wardle, under review), helping

explain how environments and genes combine to determine weight. We proposed that individuals who are less responsive to internal satiety cues by virtue of their genetic blueprint will be more likely to eat to excess when confronted by the multiple eating opportunities of the modern 'obesogenic' environment, and consequently gain more weight. Satiety responsiveness could therefore provide an important potential target for behavioural or pharmacological interventions.

NON-CODING DNA

John Mattick (Mercer, Dinger, & Mattick, 2009) and others have also demonstrated the importance of investigating the non-coding parts of the genome. Sequencing will strengthen the empirical work in this area. Not so long ago non-coding regions of the genome were thought of as an evolutionary 'junk', but now these non-coding regions are known to play an important role. For example, non-coding RNAs have been implicated in regulation of genetic expression (e.g. Shimoni et al., 2007) and small fragments of the genome that can 'move' from location to location, known as transposons, have been shown to play a role in regulating genetic expression and in creating new genes (Muotri, Marchetto, Coufal, & Gage, 2007). Just as with rare variants, earlier discussed sequencing, has a potential to illuminate sophisticated regulatory network systems and contribute to our understanding of genetic responses to environmental inputs.

ANIMAL MODELS

Working with animal models offers a variety of different approaches to enhance our capacity to learn more about pathways from genes to behaviours. Animal models have not yet been used to investigate the question of 'missing heritability'; yet they may contribute considerably. The benefits of working with mice include control over the genome, the environment, and mating. This provides opportunities for several powerful research strategies.

Recombinant inbred strains are suitable for approaches that require control over the genome and over the environment. Because each strain has a 'clonal' genome, we can test the effects of different environments on the same genome – in a setting that is conceptually similar to the MZ twin design in classical twin modelling. This approach would be particularly useful for studying gene-environment interactions.

Outbred stock would be useful for estimating the heritability of any trait under consideration because large, and known kinship networks (such as given with F1, F2 generations, siblings, and so on), increase the power of classical quantitative genetic studies.

Another animal model strategy would be to take advantage of recent work in sequencing. The widely used inbred mouse strain, C57BL/6J, was sequenced in 2002 (Wade et al., 2002), and there are now 17 strains that have been deeply sequenced on an Illumina platform (Keane et al., 2011; Yalcin et al., 2011) as part of the Mouse Genomes Project (http://www.sanger.ac.uk/resources/mouse/genomes/). Mouse genetics research has accelerated as costs of sequencing decreased and catalogues of rare structural variations - such as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), long terminal repeats (LTRs) and variable-number tandem repeats (VNTRs), - is greater than for any other mammal (Yalcin, Adams, Flint & Keane, 2012). It is a good time to consider applying GCTA methods to this burgeoning field of research. As third generation sequencing develops, with increased marker density, more accurate GCTA models will become feasible (Yalcin et al., 2012). For example, a rich catalogue of rare variants will enable calculation of an expanded relatedness matrix, offering improved estimation of genetic influence on complex traits from both common and rare variants. It will also allow us to discover if there is greater aetiological agreement of key parameter estimates between GCTA studies that use common SNPs and classical quantitative genetics. Since multiple repeated measurements can be carried out non-invasively and longitudinally for traits such as BMI, mice offer enhanced reliability and life-course characterisation. One could track, with little sample attrition, estimates of heritability through the life-course, with greater ease of precision than feasible using human samples.

CONCLUSION

In conclusion, the first three studies in this thesis suggest that a significant proportion of the 'missing heritability' is due to tiny effect sizes that are very hard to detect with current samples and methods. The results also suggest that rarer variants and non-additive genetic effects will contribute importantly to narrowing the 'missing heritability' gap. The direct comparison between the twin and molecular genetic methods enabled me to show, in the final three studies, that some of the most consistent findings from twin research can be directly replicated using DNA alone. Through this I have empirically confirmed the usefulness of the twin method and concluded that the twin method remains a useful tool for elucidating for the genetic and environmental architecture underlying complex traits.

REFERENCES

Elks, C. E., den Hoed, M., Zhao, J. H., Sharp, S. J., Wareham, N. J., Loos, R. J., & Ong, K. K. (2012). Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne),* 3, 29. doi: 10.3389/fendo.2012.00029.

Farooqi, I. S., & O'Rahilly, S. (2006). Genetics of obesity in humans. *Endocrine reviews, 27*(7), 710-718.

Girirajan, S., Rosenfeld, J. A., Coe, B. P., Parikh, S., Friedman, N., Goldstein, A., . . . Meschino, W. S. (2012). Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *New England Journal of Medicine, 367*(14), 1321-1331.

Haworth, C., Carnell, S., Meaburn, E. L., Davis, O. S. P., Plomin, R., & Wardle, J. (2008). Increasing heritability of BMI and stronger associations with the FTO gene over childhood. *Obesity, 16*(12), 2663-2668.

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., . . . Goodson, M. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature, 477*(7364), 289-294.

Llewellyn, C. H., Trzaskowski, M., van Jaarsveld, C. H. M., Plomin, R., & Wardle, J. (under review). Satiety mechanisms in genetic risk of obesity. *the Journal of the American Medical Association*.

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews Genetics, 10*(3), 155-159.

Muotri, A. R., Marchetto, M. C. N., Coufal, N. G., & Gage, F. H. (2007). The necessary junk: new functions for transposable elements. *Human molecular genetics, 16*(R2), R159-R167.

Nichols, P. L. (1984). Familial mental retardation. *Behavior Genetics, 14*(3), 161-170.

Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*.

Qi, Q., Li, Y., Chomistek, A. K., Kang, J. H., Curhan, G. C., Pasquale, L. R., . . . Qi, L. (2012). Television Watching, Leisure Time Physical Activity, and the Genetic Predisposition in Relation to Body Mass Index in Women and MenClinical Perspective. *Circulation, 126*(15), 1821-1827.

Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., & Margalit, H. (2007). Regulation of gene expression by small non-coding RNAs: a quantitative view. *Molecular systems biology, 3*(1).

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., . . . MV§gi, R. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics, 42*(11), 937-948.

Wade, C. M., Kulbokas, E. J., Kirby, A. W., Zody, M. C., Mullikin, J. C., Lander, E. S., . . . Daly, M. J. (2002). The mosaic structure of variation in the laboratory mouse genome. *Nature, 420*(6915), 574-578. doi: 10.1038/nature01252

Wheeler, E., Huang, N., Bochukova, E. G., Keogh, J. M., Lindsay, S., Garg, S., . . . Farooqi, I. S. (2013). Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nature Genetics, 45*(5), 513-517. doi: 10.1038/ng.2607

Yalcin, B., Adams, D. J., Flint, J., & Keane, T. M. (2012). Next-generation sequencing of experimental mouse strains. *Mammalian Genome, 23*(9-10), 490-498.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T. M., Gan, X., . . . Bhomra, A. (2011). Sequence-based characterization of structural variation in the mouse genome. *Nature, 477*(7364), 326-329.

Zahnleiter, D., Uebe, S., Ekici, A. B., Hoyer, J., Wiesener, A., Wieczorek, D., . . . Rauch, A. (2013). Rare Copy Number Variants Are a Common Cause of Short Stature. *Plos Genetics, 9*(3), e1003365.

APPENDICES

APPENDIX 1. SUPPLEMENTARY TEXT MATERIALS FOR CHAPTER 2

GENOTYPING PROTOCOL

DNA was extracted from buccal cheek swabs and sent toAffymetrix, Santa Clara, California, USA. In total, 3,665 samples were successfully hybridized to AffymetrixGeneChip 6.0 SNP genotyping arrays (http://www.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf) using experimental protocols recommended by the manufacturer (Affymetrix Inc., Santa Clara, CA). The raw image data from the arrays were normalized and pre-processed at the Wellcome Trust Sanger Institute, Hinxton, UK for genotyping as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/) according to the manufacturer's guidelines (http://www.affymetrix.com/support/downloads/manuals/genomewidesnp6_manual.pdf). Genotypes for the Affymetrix arrays were called using CHIAMO (https://mathgen.stats.ox.ac.uk/genetics_software/chiamo/chiamo.html).

Where there was a sufficient quantity of DNA, samples were also re-genotyped on a panel of 30 SNPs (including 26 autosomal SNPs present on the Affymetrix array, and 4 SNPs on the X chromosome to verify gender) using the SequenomiPlex Gold assay (Sequenom Inc., San Diego, CA).

QUALITY CONTROL: SAMPLES

We identified and removed samples whose genome-wide patterns of diversity differed from those of the collection at large, interpreting these differences as likely to be due to biases or artifacts.

Outlying individuals were identified on the basis of call rate, heterozygosity, relatedness and ancestry using a Bayesian clustering approach [1].

To obtain a set of putatively unrelated individuals we used a hidden Markov model (HMM) to infer identify by descent along the genome between pairs of individuals. Amongst pairs of closely related individuals, we excluded the member of the pair with the lowest call rate, iteratively repeating this procedure to obtain a set of individuals with pairwise identity by decent less than 5% [1].

Of the individuals genotyped, samples were excluded because of low call rate or heterozygosity outliers (377), unusual hybridization intensity (9), atypical population ancestry (59), sample duplication or relatedness to other sample members (83), and gender mismatches (13). In addition, 54 samples were excluded because fewer than 90% of genotypes were called identically on the genome-wide array and Sequenom panel. The remaining samples were consistent with previous genotyping. In total, 513 samples were excluded by these quality control criteria. The remaining sample of 3,152 individuals included 1,446 males and 1,706 females.

QUALITY CONTROL: SNPS

A measure of information for the allele frequency at each of 932,533 called SNPs was calculated using SNPTEST version 2.1.1 [2]. Autosomal SNPs were excluded if this information measure was below 0.975, if the minor allele frequency was less than 1%, if greater than 2% of genotype data were missing, or if the Hardy Weinberg *p*-value was lower than 10^{-20}. Association between the SNP and the plate on which samples were genotyped was calculated; SNPs with a plate effect *p*-value less than 10^{-6} were also excluded. In addition, SNPs were manually filtered for call quality by visual

inspection of the hybridization intensity plots using EVOKER software (http://sourceforge.net/projects/evoker/). The above filters removed 26 % of the SNPs, leaving 690,943 autosomal SNPs for further analysis.

STATISTICAL ANALYSIS

Imputation was carried out using the IMPUTE version 2 software [3] on the genotype data after application of quality control procedures, using a two-stage approach with both a haploid reference panel and a diploid reference panel. For the haploid reference panel we used HapMap phase II and III SNP data on the 120 unrelated CEU trios. 5,175 WTCCC2 controls were genotyped on both Affymetrix 6.0 and Illumina Human1.2M-Duo arrays (Illumina Inc., La Jolla, CA), and these were used for the diploid reference panel. SNPs were retained for analysis if they were genotyped using the Affymetrix 6.0 array, if they were genotyped using the Illumina Human1.2M-Duo array and obtained an information score ≥ 0.90, or if they were imputed and obtained an information score ≥ 0.98. Using these criteria, 1,033,462 imputed SNPs were retained giving a total of 1,724,405 SNPs used in the association analyses.

We performed Principal Component Analysis on a subset of 105,556 autosomal SNPs remaining after applying our quality control criteria, after pruning to remove SNPs in high linkage disequilibrium ($r^2 > 0.2$) and excluding high linkage disequilibrium genomic regions so as to ensure that only genome-wide effects were detected (Fellay et al., 2007). Application of the Tracy-Widom test [4] indicated that eight principal components were significant using a threshold of $p < 0.05$. Phenotype scores were normalized by transforming the ranked data to the quantiles of a standard normal distribution using the van der Waerden transformation [5]. Linear regression analysis was run for

each autosomal SNP using a frequentist additive model that accounts for uncertain genotype data, as implemented in SNPTEST version 2.1.1 [2]. Sex and the first eight principal components of the genotype data were included as covariates in the regression model.

SOM REFERENCES

1. Barrett JC, Lee JC, Lees CW, Prescott NJ, Anderson CA, et al. (2009) Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nat Genet 41: 1330-1334. Epub 2009 Nov 1315.

2. Marchini J, Howie BN, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies via imputation of genotypes. Nature Genetics 38: 906-913.

3. Howie BN, Donnelly P, Marchini J (2009) A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. Plos Genetics 5: e1000529.

4. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. Plos Genetics 2: e190.

5. van der Waerden BL (1952) Order tests for the two-sample problem and their power. Indagationes Mathematicae: 453-458.

APPENDIX 2. SUPPLEMENTARY ONLINE MATERIALS FOR CHAPTER 3

eTABLE 1. GCTA AND TWIN GENETIC ESTIMATES FOR BEHAVIORAL PROBLEMS SCALES. BOLD FONT INDICATES RESULTS PRESENTED IN FIGURE 2.

A. Child self-reports	**Twin heritability**							**GCTA heritability**				
	A	***SE***	***C***	***SE***	***E***	***SE***	***n/pairs***	***V(G)***	***SE***	***V(e)***	***SE***	***n***
MFQ (depressive symptoms)	**0.38**	**0.06**	**0.09**	**0.05**	**0.53**	**0.02**	**2683**	**0.00**	**0.12**	**1.00**	**0.12**	**2698**
SDQ Behavior Problems composite	**0.44**	**0.06**	**0.09**	**0.05**	**0.47**	**0.02**	**2668**	**0.00**	**0.12**	**1.00**	**0.13**	**2690**
Anxiety	**0.41**	**0.03**	**0.00**	**0.01**	**0.59**	**0.02**	**2668**	**0.02**	**0.12**	**0.99**	**0.13**	**2687**
Conduct	**0.37**	**0.06**	**0.06**	**0.04**	**0.57**	**0.02**	**2670**	**0.00**	**0.12**	**1.00**	**0.12**	**2690**
Hyperactivity	**0.46**	**0.03**	**0.00**	**0.01**	**0.54**	**0.02**	**2672**	**0.00**	**0.12**	**1.00**	**0.12**	**2687**
Peer Problems	**0.40**	**0.05**	**0.02**	**0.03**	**0.57**	**0.02**	**2674**	**0.00**	**0.12**	**1.00**	**0.12**	**2692**

B. Parent ratings	Twin heritability							GCTA heritability				
	A	*SE*	*C*	*SE*	*E*	*SE*	*n/pairs*	*V(G)*	*SE*	*V(e)*	*SE*	*n*
Conners ADHD composite	**0.80**	**0.03**	**0.05**	**0.03**	**0.15**	**0.01**	**2686**	**0.00**	**0.12**	**1.00**	**0.12**	**2692**
Hyperactivity-Impulsivity	0.79	0.04	0.09	0.04	0.12	0.01	2685	0.06	0.12	0.93	0.12	2688
Inattention	0.79	0.01	0.00	0.01	0.21	0.01	2687	0.00	0.12	1.00	0.12	2693
APSD psychopathic symptoms composite	**0.49**	**0.03**	**0.35**	**0.03**	**0.16**	**0.01**	**2694**	**0.00**	**0.12**	**1.00**	**0.12**	**2699**
Callous-Unemotional	0.31	0.03	0.54	0.02	0.15	0.01	2694	0.02	0.12	0.98	0.12	2700
Impulsivity	0.66	0.04	0.16	0.04	0.19	0.01	2687	0.00	0.12	0.99	0.12	2697
Narcissism total	0.63	0.04	0.16	0.04	0.22	0.01	2695	0.00	0.12	1.00	0.12	2700
CAST autistic symptoms composite	**0.73**	**0.04**	**0.06**	**0.03**	**0.22**	**0.01**	**2688**	**0.09**	**0.12**	**0.91**	**0.12**	**2694**
Communication	0.76	0.03	0.01	0.02	0.23	0.01	2689	0.00	0.12	1.00	0.12	2692
Non-Social	0.72	0.01	0.00	0.01	0.28	0.01	2689	0.00	0.12	1.00	0.12	2692
Social	0.71	0.02	0.00	0.01	0.29	0.01	2687	0.06	0.12	0.94	0.12	2693
MFQ (depressive symptoms)	**0.71**	**0.04**	**0.06**	**0.03**	**0.24**	**0.01**	**2680**	**0.00**	**0.12**	**1.00**	**0.12**	**2688**
SDQ Behavior Problems composite	**0.60**	**0.04**	**0.21**	**0.03**	**0.19**	**0.01**	**2687**	**0.00**	**0.12**	**1.00**	**0.12**	**2692**
Anxiety	0.61	0.04	0.02	0.03	0.38	0.02	2683	0.00	0.12	1.00	0.12	2687
Conduct total	0.55	0.04	0.22	0.04	0.23	0.01	2685	0.00	0.12	1.00	0.12	2691
Hyperactivity	0.78	0.01	0.00	0.01	0.22	0.01	2687	0.00	0.12	1.00	0.12	2691
Peer Problems	0.78	0.01	0.00	0.01	0.22	0.01	2685	0.16	0.12	0.84	0.12	2690

C. Teacher ratings	Twin heritability							GCTA heritability				
	A	*SE*	*C*	*SE*	*E*	*SE*	*n/pairs*	*V(G)*	*SE*	*V(e)*	*SE*	*n*
APSD psychopathic symptoms composite	**0.61**	**0.05**	**0.02**	**0.03**	**0.37**	**0.02**	**1901**	**0.15**	**0.16**	**0.85**	**0.16**	**2129**
Callous-Unemotional	0.32	0.08	0.12	0.06	0.56	0.03	1891	0.00	0.16	1.00	0.16	2125
Impulsivity	0.63	0.02	0.00	0.01	0.37	0.02	1898	0.24	0.16	0.76	0.16	2120
Narcissism	0.65	0.02	0.00	0.01	0.35	0.02	1904	0.50	0.16	0.50	0.15	2128
CAST composite	**0.48**	**0.04**	**0.00**	**0.02**	**0.52**	**0.02**	**1896**	**0.00**	**0.16**	**1.00**	**0.16**	**2120**
Communication	0.50	0.03	0.00	0.01	0.50	0.02	1899	0.00	0.15	1.00	0.16	2121
Non-Social	0.47	0.05	0.01	0.03	0.52	0.03	1783	0.00	0.16	1.00	0.16	2034
Social total	0.49	0.03	0.00	0.02	0.51	0.02	1886	0.00	0.16	1.00	0.16	2117
SDQ Behavior Problems composite	**0.59**	**0.03**	**0.00**	**0.02**	**0.41**	**0.02**	**1919**	**0.11**	**0.15**	**0.90**	**0.15**	**2137**
Anxiety	0.52	0.05	0.02	0.03	0.46	0.02	1912	0.11	0.15	0.88	0.15	2135
Conduct	0.54	0.02	0.00	0.01	0.46	0.02	1921	0.26	0.15	0.73	0.15	2137
Hyperactivity	0.57	0.03	0.00	0.01	0.43	0.02	1925	0.05	0.15	0.95	0.15	2138
Peer Problems	0.54	0.04	0.00	0.02	0.46	0.02	1918	0.00	0.15	1.00	0.16	2139

Annotation: MFQ – Moods and Feelings Questionnaire; SDQ – Strengths and Difficulties Questionnaire; ADHD – Attention Deficit Hyperactivity Disorder; APSD – Antisocial Process Screening Device; CAST – Childhood Asperger Syndrome Test; A – heritability estimate; C – shared environment estimate; E – unique environment estimate; SE – standard error; N/pairs – number of twin pairs with available data; V(G) – variance explained by additive genetic factors; n – number of unrelated individuals with available data.

Figure S1a Histograms of untransformed composite scales.

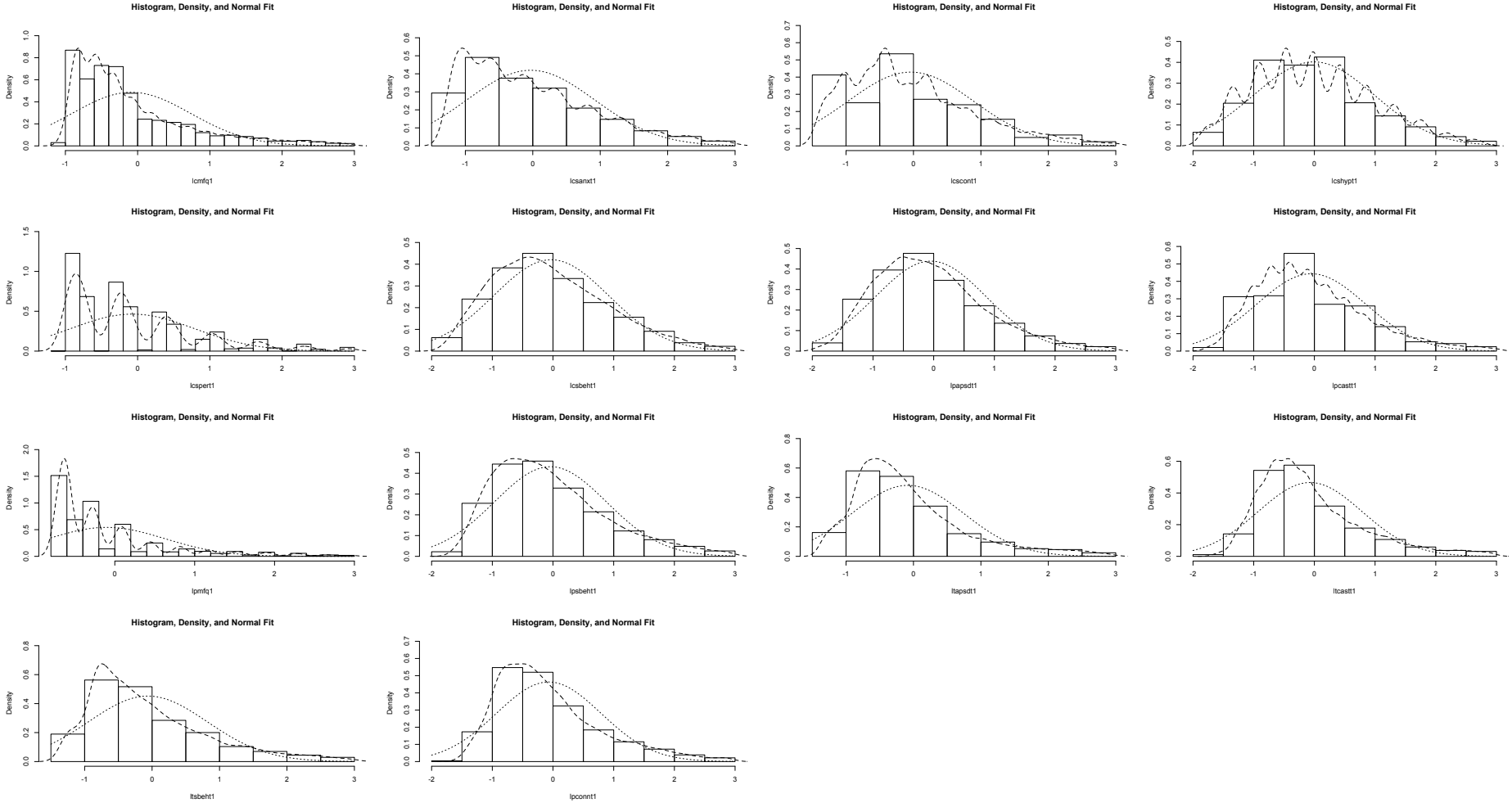
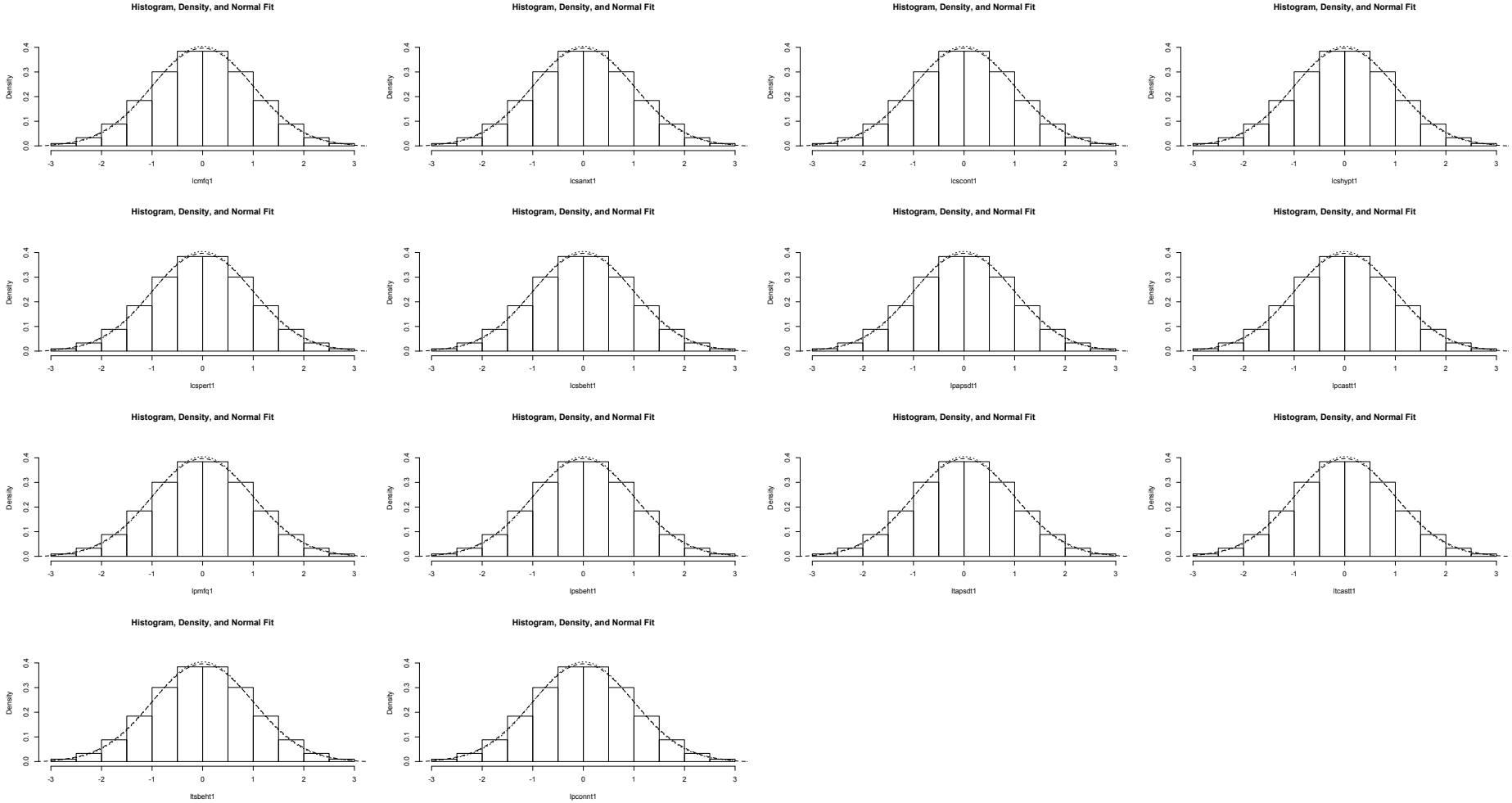


Figure S2b Histograms of composite quantile normalized scales.



Annotation: 'lcmfq1' – child self-rated MFQ; 'lcsanxt1' – child self-rated SDQ Anxiety; 'lcscont1' – child self-rated SDQ Conduct; 'lcshypt1' – child self-rated SDQ Hyperactivity; 'lcspert1' – child self-rated SDQ Peer Problems; 'lcsbeht1' – child self-rated SDQ composite; 'lpapsdt1' – parent-rated APSD composite; 'lpcstt1' – parent-rated CAST composite; 'lpmfq1' – parent-rated MFQ composite; 'lpsbeht1' – parent-rated SDQ composite; 'ltapsdt1' – teacher-rated APSD composite; 'ltcastt1' – teacher-rated CAST composite; 'ltsbeht1' – teacher-rated SDQ composite; 'lpconnt1' – parent-rated CONNERS ADHD composite;

Footnote: As seen in Figure S1a, some of the composite scales are skewed, as is typical of behavior problem scales. However, Figure S1b shows that van der Wearden transformation (see Methods) normalizes the scales. It is noteworthy that despite the considerable transformation of the distributions, the correlation between the untransformed and transformed scales are high (.0.90, 0.96, 0.96, 0.99, 0.92, 0.98, 0.98, 0.97, 0.86, 0.97, 0.94, 0.95, 0.94, 0.95, respectively), indicating that the transformation did not drastically disrupt the rank-order structure of the data. In the Results section of the text, we presented results for the transformed scales; however, as a further check on the effect of non-normality, we also compared GCTA and twin point estimates of heritability for the transformed and untransformed scales. We found that the largest GCTA heritability difference was 0.04 and the largest twin heritability difference was also 0.04.

APPENDIX 3. SUPPLEMENTARY ONLINE MATERIALS FOR CHAPTER 5

SUPPLEMENTARY TABLE 1. BIVARIATE GENOME-WIDE COMPLEX TRAIT ANALYSIS (GCTA) RESULTS (WITH STANDARD ERRORS) FOR BMI-SDS SCORES BETWEEN AGES 4 AND 10 YEARS

variables	A						E				Vp_tr1	Vp_tr2	n_tr1*	n_tr2*
	V(G)_tr1	V(G)_tr2	C(G)_tr12	V(G) /Vp_tr1	V(G) /Vp_tr2	r_G	V(e)_tr1	V(e)_tr2	C(e)_tr12	r_E**				
BMISDS 4 to 11	.20(.21)	.29(.14)	.16(.13)	.20(.21)	.29(.14)	.66(.48)	.81(.21)	.71(.14)	.25(.13)	.34(.15)	1.00(.04)	1.00(.03)	1419	2268

Annotation: V(G) – variance explained by genetic factors for trait 1 and trait 2 (tr1, tr2); C(G) – covariance between trait 1 and 2 explained by genetic factors; V(e) – residual variance for trait 1 and trait 2; C(e) – residual covariance between trait 1 and trait 2; Vp – phenotypic variance for trait 1 and trait 2; V(G) / Vp – proportion of the phenotypic variance explained by genetic factors for trait 1 and trait 2; r_G – genetic correlation between trait 1 and trait 2; logL – log likelihood estimation of the model; n – number of individuals with data for both trait 1 and trait 2; values in parentheses are standard errors.

*GCTA incorporates full-information maximum likelihood that uses the full sample of 2556 individuals with data on trait 1 or trait 2. However, the variance estimates for each trait are based on individuals with data for that trait, the last two columns are sample sizes with data present at each age.

** The current version of GCTA does not report the environmental correlation or its standard error. The environmental correlation was derived here from the GCTA estimates using the following algorithm: C(e)_tr12 / (√V(e)_tr1 * √V(e)_tr2), whereas the standard error was calculated using: Var(re) = re * re * (VarVe1/(4*Ve1*Ve1) + VarVe2/(4*Ve2*Ve2) + VarCe/(Ce*Ce) + CovVe1Ve2/(2*Ve1*Ve2) - CovVe1Ce/(Ve1*Ce) - CovVe2Ce/(Ve2*Ce)); SE(re) = sqrt[Var(re)], where re is the environmental correlation, Ve1 is the residual variance for trait 1, Ce is the residual covariance between two traits, VarVe1 is the sampling variance for Ve1 (residual variance for trait 1), VarCe is the sampling variance for Ce, CovVe1Ve2 is the sampling covariance between Ve1 and Ve2, and CovVe1Ce is the sampling covariance between Ve1 and Ce.

SUPPLEMENTARY TABLE 2. BIVARIATE TWIN MODEL-FITTING RESULTS (WITH STANDARD ERRORS) FOR BMI-SDS BETWEEN AGES 4 AND 10 YEARS

variables	A				C				E				n/pairs*
	V(G)_tr1	V(G)_tr2	C(G)_tr12	r_G	V(c)_tr1	V(c)_tr2	C(c)_tr12	r_C	V(e)_tr1	V(e)_tr2	C(e)_tr12	r_E	
BMISDS 4 to 11	.43(.04)	.82(.04)	.34(.04)	.58(.05)	.41(.04)	.06(.04)	.01(.04)	.08(.51)	.16(.01)	.12(.01)	.04(.01)	.31(.04)	3104

Annotation: V(G) – proportion of the variance explained by genetic factors for trait 1 and trait 2 (tr1, tr2); C(G) – proportion of the covariance between trait 1 and 2 explained by genetic factors; V(c) – proportion of the variance explained by shared environment for trait 1 and trait 2 (tr1, tr2); C(c) – proportion of the covariance between trait 1 and 2 explained by shared environment; V(e) – proportion of the variance explained by non-shared environment for trait 1 and trait 2 (tr1, tr2); C(e) – proportion of the covariance between trait 1 and 2 explained by non-shared environment; rG – genetic correlation; rC – correlation of shared environmental factors; rE – correlation of non-shared environmental factors; n – number of twin pairs with data for both trait 1 and trait 2; values in parentheses are standard errors.

*OpenMx twin model-fitting incorporates full-information maximum likelihood that uses the full sample where at least one sibling has available data.

SUPPLEMENTARY FIGURE 1. SAMPLING DISTRIBUTION FOR R^2 AT AGE 4 (A) AT AGE 10 (B) AND FOR DIFFERENCE IN R^2 BETWEEN THE AGES 4 AND 10